

**Técnicas de Minería de Datos para la Predicción de Fallas en las Líneas de
Embotellado en Bavaria Bucaramanga**

Yefferson Espinosa Pacheco

Trabajo de Grado para Optar por el Título de Ingeniero Industrial

Director

Henry Lamos Diaz

PhD. en Física Matemáticas

Universidad Industrial de Santander

Facultad de Ingenierías Físico-Mecánicas

Escuela de Estudios Industriales y Empresariales

Bucaramanga

2018

AGRADECIMIENTOS

Primero que todo a Papá Dios por que siempre ha sido generoso conmigo.

A la Universidad Industrial de Santander, especialmente a mi escuela de estudios industriales y empresariales por dar su granito de arena a mi formación profesional.

A Henry Lamos Diaz, director, por su sublime paciencia, transferencia de conocimiento y confianza en el desarrollo de este proyecto.

A la Gerencia de Embotellado de Bavaria S.A sede Bucaramanga, por impulsar y confiar en este trabajo.

Al grupo de investigación OPALO por la disponibilidad de recursos para la ejecución de los objetivos.

A todos mis amigos, compañeros y profesores quienes han acompañado y aportado valor a mi carrera profesional.

Tabla de Contenido

Introducción	17
Tabla de cumplimiento de objetivos	19
1. Planteamiento del problema.....	20
2. Justificación.....	21
3. Objetivos	23
3.1 Objetivo general.....	23
3.2 Objetivos específicos	23
4. Caracterización del sistema productivo.....	24
4.1 Bavaria S.A.....	24
4.2 Cervecería de Bucaramanga	25
4.3 Área de Embotellado de Bavaria Bucaramanga.....	26
4.3.1 Estructura organizacional.....	26
4.3.2 Descripción del proceso de envasado	26
4.3.3 Rendimiento de una línea.....	35
5. Marco teórico	36
5.1 Análisis preliminar de literatura.....	36

TÉCNICAS MD PREDICCIÓN FALLAS EMBOTELLADO BAVARIA	6
5.2 Conceptos básicos.....	38
5.3 Formas de mantenimiento.....	40
5.4 Descubrimiento de conocimiento en bases de datos.....	42
5.5 Análisis de Componentes Principales	46
5.6 Regresión Lineal Múltiple	48
6. Proceso de descubrimiento de conocimiento en bases de datos	50
6.1 Definición de objetivos.....	50
6.2 Preparación de los datos.....	50
6.3 Procesamiento de los datos	51
6.3.1 Variables eliminadas.....	51
6.3.2 Variables modificadas.....	51
6.3.3 Variables creadas.	51
6.3.4 Definición de variables definitivas	53
6.4 Minería de datos.....	54
6.4.1 Análisis previo de los datos.	55
6.4.2 Paso a paso del análisis de componentes principales.....	57
6.4.3 Paso a paso regresión lineal múltiple.....	58
7. Resultados	63
7.1 Resultados análisis previo de los datos	63

TÉCNICAS MD PREDICCIÓN FALLAS EMBOTELLADO BAVARIA	7
7.2 Resultados análisis de componentes principales.....	74
7.3 Resultados regresión lineal múltiple.....	84
8. Modelo estándar.....	90
9. Conclusiones.....	91
10. Recomendaciones.....	92
Referencias Bibliográficas.....	93

Lista de Tablas

Tabla 1. <i>Cumplimiento de los objetivos del proyecto</i>	19
Tabla 2. <i>Variables eliminadas de Zenvase2015-2017 línea 2</i>	52
Tabla 3. <i>Variables modificadas</i>	52
Tabla 4. <i>Variables creadas</i>	52
Tabla 5. <i>Lista de máquinas críticas</i>	65
Tabla 6. <i>Meses críticos de cada máquina</i>	69
Tabla 7. <i>Componentes principales seleccionados</i>	78
Tabla 8. <i>Aplicación de modelo matemático para datos del año 2018</i>	89

Lista de Figuras

<i>Figura 1.</i> Logotipo Bavaria y ABInBev.	24
<i>Figura 2.</i> Ubicación geográfica Cervecería Bucaramanga.	25
<i>Figura 3.</i> Instalaciones área de embotellado	26
<i>Figura 4.</i> Estructura organizacional equipo embotellado.....	27
<i>Figura 5.</i> Descripción del proceso de embotellado	27
<i>Figura 6.</i> Depaletizadora	28
<i>Figura 7.</i> Desempacadora.....	29
<i>Figura 8.</i> Lavadora de botellas	29
<i>Figura 9.</i> Inspector de botellas vacías	30
<i>Figura 10.</i> Envasadora y coronador.....	31
<i>Figura 11.</i> Pasteurizador.....	31
<i>Figura 12.</i> Etiquetadora	32
<i>Figura 13.</i> Inspector de producto terminado	32
<i>Figura 14.</i> Lavadora de cajas.....	33
<i>Figura 15.</i> Empacadora	33
<i>Figura 16.</i> Inspector de caja llena.....	34
<i>Figura 17.</i> Paletizadora.....	34
<i>Figura 18.</i> Gráfico en V.....	35
<i>Figura 19.</i> Relación datos, registro, archivo y bases de datos.....	38
<i>Figura 20.</i> Diagrama de mantenimiento	41
<i>Figura 21.</i> Proceso de Descubrimiento de Conocimiento en Bases de datos.....	43
<i>Figura 22.</i> Metodología SMART	43

TÉCNICAS MD PREDICCIÓN FALLAS EMBOTELLADO BAVARIA	10
<i>Figura 23.</i> Análisis exploratorio.....	55
<i>Figura 24.</i> ANOVA de un factor.....	55
<i>Figura 25.</i> Opciones del ANOVA de un factor.....	56
<i>Figura 26.</i> Resultados del ANOVA de un factor.....	56
<i>Figura 27.</i> Análisis en PCA.....	57
<i>Figura 28.</i> Código para gráficas en PCA.....	58
<i>Figura 29.</i> Análisis de análisis de regresión múltiple.....	59
<i>Figura 30.</i> Selección de variables en regresión múltiple.....	59
<i>Figura 31.</i> Cálculo de variable predictora.....	60
<i>Figura 32.</i> Análisis de prueba T para muestras relacionadas.....	60
<i>Figura 33.</i> Creación de variables dummy.....	61
<i>Figura 34.</i> Producto vs media de tiempo total.....	64
<i>Figura 35.</i> Prueba T de muestras independientes para pilsen y poker.....	64
<i>Figura 36.</i> Gráfica producto vs media de tiempo perdido en lavadora de botellas.....	65
<i>Figura 37.</i> Gráfica producto vs media de tiempo perdido en transportado de botellas.....	66
<i>Figura 38.</i> Gráfica producto vs media de tiempo perdido en envasadora.....	66
<i>Figura 39.</i> Gráfica producto vs media de tiempo perdido en etiquetadora.....	67
<i>Figura 40.</i> ANOVA de tiempo perdido por meses.....	67
<i>Figura 41.</i> Gráfica mes vs media de tiempo total perdido.....	68
<i>Figura 42.</i> Gráfica mes vs media de tiempo perdido en lavadora de botellas.....	69
<i>Figura 43.</i> Gráfica mes vs media de tiempo perdido en transportador de botellas.....	70
<i>Figura 44.</i> Gráfica mes vs media de tiempo perdido en envasadora.....	70
<i>Figura 45.</i> Gráfica mes vs media de tiempo perdido en etiquetadora.....	71

TÉCNICAS MD PREDICCIÓN FALLAS EMBOTELLADO BAVARIA	11
<i>Figura 46.</i> ANOVA de tiempo perdido por semana.....	71
<i>Figura 47.</i> Gráfica semanas vs media de tiempo total perdido	72
<i>Figura 48.</i> Gráfica semana vs media de tiempo perdido en lavadora de botellas	72
<i>Figura 49.</i> Gráfica semana vs media de tiempo perdido en transportador de botellas.....	73
<i>Figura 50.</i> Gráfica semana vs media de tiempo perdido en envasadora	73
<i>Figura 51.</i> Gráfica semana vs media de tiempo perdido en etiquetadora	74
<i>Figura 52.</i> Prueba KMO y Bartlett.....	74
<i>Figura 53.</i> Gráfico de sedimentación	75
<i>Figura 54.</i> Componentes principales.....	76
<i>Figura 55.</i> Distribución de las componentes en el layout actual de la línea de producción.....	79
<i>Figura 56.</i> Ubicación de máquinas respecto a PC1 y PC2.....	79
<i>Figura 57.</i> Coordenadas de los individuos respecto al PC1 y PC2	80
<i>Figura 58.</i> Coordenadas de los productos de cada individuo respecto a CP1 y CP2.....	81
<i>Figura 59.</i> Ubicación de máquinas respecto a CP1 y CP3.....	82
<i>Figura 60.</i> Coordenadas de los individuos respecto al CP1 y CP3	82
<i>Figura 61.</i> Ubicación de máquinas respecto a CP2 y CP3	83
<i>Figura 62.</i> Coordenadas de los individuos respecto al CP1 y CP3	83
<i>Figura 63.</i> Coeficientes beta de las variables predictoras	84
<i>Figura 64.</i> Resumen de modelo de regresión lineal múltiple.....	84
<i>Figura 65.</i> ANOVA de los modelos de regresión	85
<i>Figura 66.</i> Coeficientes de los modelos de regresión.....	85
<i>Figura 67.</i> ANOVA del modelo de regresión 2	87
<i>Figura 68.</i> Coeficientes para el modelo 2.....	87

TÉCNICAS MD PREDICCIÓN FALLAS EMBOTELLADO BAVARIA	12
<i>Figura 69.</i> Estadísticos para las muestras relacionadas: variable real vs variable nueva.....	88
<i>Figura 70.</i> Correlaciones de las muestras relacionadas: variable real vs variable nueva.....	88
<i>Figura 71.</i> Prueba T para las muestras relacionadas: variable real vs variable nueva	88
<i>Figura 72.</i> Modelo estándar.....	90

Lista de Apéndices

**(Ver apéndices adjuntos en el CD y pueden visualizarlos en la Base de Datos de la
Biblioteca UIS)**

Apéndice A. Descripción del proceso de embotellado

Apéndice B. Datos Zenvase2015-2017 línea 2

Apéndice C. Datos Ztiemps2015-2017 línea 2

Apéndice D. Datos regresión lineal múltiple

Apéndice E. Artículo de carácter publicable.

Glosario

Minería de datos: proceso de descubrimiento de conocimiento sobre almacenes de datos complejos mediante la extracción oculta y potencialmente útil en forma de patrones globales y relaciones estructurales implícitas entre datos.

Descubrimiento de conocimiento en bases de datos: proceso para identificar patrones novedosos, útiles y potencialmente entendibles entro de un gran número de datos.

Análisis de componentes principales: es uno de los métodos de análisis más difundidos, que permite la estructuración de un conjunto de datos multivariado obtenidos de una población, cuya distribución de probabilidades no necesita ser conocida.

Regresión lineal múltiple: es una técnica estadística que puede utilizarse para analizar la relación entre una única variable criterio y varias variables independientes.

Productos: mercancía producida con el uso de maquinaria y tecnología en una fábrica.

Variables dummy: variables que toma que toma el valor de 0 y 1 cuando indica la ausencia o presencia de algún efecto categórico que afecta el resultado final.

RESUMEN

TÍTULO

TÉCNICAS DE MINERÍA DE DATOS PARA LA PREDICCIÓN DE FALLAS EN LAS LÍNEAS DE EMBOTELLADO EN BAVARIA BUCARAMANGA¹

AUTOR

ESPINOSA PACHECO, Yefferson²

PALABRAS CLAVES

Minería de datos, Descubrimiento de conocimiento en bases de datos, Análisis de componentes principales, Regresión lineal múltiple, Productos, Variables dummy.

DESCRIPCIÓN

En búsqueda de estrategias competitivas e innovadora que ayuden a industrias manufactureras a disminuir el tiempo perdido durante sus procesos, identificando puntos críticos y evidenciando tendencias que no son visibles a simple vista; surge el concepto de Descubrimiento de Conocimiento en Bases de Datos especialmente la etapa de minería de datos. En la presente investigación se muestra la aplicación de la minería de datos en una de las empresas más importante del país que pertenece al sector de bebidas. Se usan las herramientas de análisis exploratorio, análisis de componente principales (PCA) y regresión lineal múltiple con el propósito de generar conocimiento para la creación de planes estratégicos de mantenimiento en base a los focos críticos identificados y la propuesta de un modelo que permita predecir la cantidad de tiempo que se perderá durante un ciclo de producción. Se utilizan los programas estadísticos IBM SPSS Statistics 23, RStudio y Minitab para la aplicación de las técnicas. Luego de la evaluación, se define el modelo de predicción de tiempo total perdido y los grupos de máquinas significativos que se generan del PCA para una mejora toma de decisiones por parte de la gerencia, cuyo resultado sea reflejado en las eficiencias y la gestión de activos de la empresa.

¹ Trabajo de grado

² Facultad de Ingenierías Físico-Mecánicas, Escuela de Estudios Industriales y Empresariales. Director PhD Henry Lamos Diaz

ABSTRACT**TITLE**

DATA MINING TECHNIQUES USED FOR THE PREDICTION OF FAILURE IN THE BOTTLING LINES OF BAVARIA BUCARAMANGA¹

AUTHOR

ESPINOSA PACHECO, Yefferson²

KEY WORDS

Data mining, Knowledge Discovery in database, Principal components analysis, Multiple linear regression, Products, dummy variables.

DESCRIPTION

In search of competitive and innovative strategies that will help manufacturing industries to decrease time during their processes, identifying critical points and observing tendencies that are not always visible at first sight; the concept of Knowledge Discovery in Databases appears, specially the Mining Data phase of the process. In the following investigation the application of Mining Data will be shown in one of the most important company of the country, which is involved in the beverage sector. Tools such as exploratory analysis, principal component analysis and multiple linear regression are used to generate the knowledge to create strategic maintenance plans based in the critical points identified and the proposition of a model that will allow the prediction of the amount of time it will be lost during the production cycle. To apply the techniques, statistical programs are used such as IBM SPSS Statistics 23, RStudio and Minitab. After the evaluation phase, the lost time prediction model and the significant machine groups are defined using the PCA to improve the decisions taken by management, which results are reflected in the efficiency and the assets management of the company.

¹ Work degree

² Faculty of Physico Mechanical Engineering, School of Industrial and Business Studies. Director PhD Henry Lamos Diaz

Introducción

La minería de datos es definida por (Kopanakis & Theodoulidis, 2003) como el proceso de descubrimiento de conocimiento sobre almacenes de datos complejos mediante la extracción oculta y potencialmente útil en forma de patrones globales y relaciones estructurales implícitas entre datos. Otros como (Witten, Frank, Hall, & Pal, 2017) apuntan que la minería de datos como aquel proceso en que se extrae conocimiento útil y comprensible, previamente desconocido, y a partir de grandes conjuntos de datos almacenados en distintos formatos.

En el panorama nacional, durante el mes de septiembre de 2017 la producción real del sector manufacturero disminuyó -1.9% con relación a igual mes de 2016, cuando registró un aumento de 4.5%. Las industrias de elaboración de bebidas presentaron un decrecimiento de 6.3% y le restó 0.7 puntos porcentuales a la variación de la industria manufacturera. Sin embargo, el sector de bebidas en los últimos doce (12) meses a septiembre de 2017, creció 1.3% por el buen comportamiento durante el 2016 (Isabel Alonso & Ibañez, 2017).

Una de las variables influyentes en este decrecimiento tiene que ver a la caída de la demanda interna por consecuencia de factores climáticos (fuertes lluvias y bajas temperaturas), que disminuyeron la disposición de adquirir bebidas. Por el contrario, según la revista (Dinero, 2017) en su informe “Así se ve el panorama de Colombia en 2018” un factor que puede influir favorablemente es el mundial de Rusia en el que Colombia participará, donde es probable que el ambiente general de los consumidores repunte.

Adicional a los pronósticos para el año 2018, el sector de bebidas registra nuevos productos, renovación de portafolio, inversiones en modernización de plantas existentes (adaptación de línea

2 de producción para elaborar cerveza Costeñita en Bavaria Bucaramanga) e incluso nuevas fábricas consolidadas y en expansión. Añadiendo, la minería de datos es un método innovador cuyo objetivo es aprovechar la información ya existente a fin de mejorar procesos, diseñar programas de mantenimiento preventivo, mejorar el rendimiento de la inversión y optimizar el uso de recursos.

En continuidad con el uso de estrategias innovadoras que generen soluciones creativas, acertadas y preventivas, se hace necesaria la utilización de técnicas de minería de datos que ayuden a encontrar desviaciones en el proceso de envasado con el fin de identificar fallas frecuentes, creación de programas de mantenimiento preventivo y aumentar la confiabilidad operacional.

La detención de cualquier proceso productivo, sin importar su naturaleza, trae una serie de costos asociados, por lo tanto, se requiere que los equipos operen en forma continua para garantizar una máxima producción, sin embargo, su naturaleza física hace que con el paso del tiempo sus elementos se deterioren, haciendo necesario detener con cierta frecuencia las operaciones que llevan a cabo para efectuar labores de mantenimiento que permitan prolongar su vida útil. De tal forma, se hace necesario organizar los periodos en los cuales se llevarán a cabo las actividades de mantenimiento para garantizar que el costo incurrido al paralizar el proceso sea mínimo.

En búsqueda de mitigar paradas en los procesos, las grandes empresas de la industria de bebidas tratan de aplicar técnicas y/o herramientas que permitan orientar eficientemente el proceso de toma de decisiones asociado a la gestión de mantenimiento. La aplicación de esta técnica en estas empresas de gran magnitud resulta fundamental ya que permiten realizar análisis rigurosos a los registros de fallas ocurridos durante los procesos de producción.

Es importante conocer la forma como se relacionan las variables que integran el registro de fallas, ya que a partir de allí se pueden identificar situaciones particulares que no son apreciables a simple vista y con ello, detectar patrones que permitan incrementar la confiabilidad de los equipos. Por tal motivo, la presente investigación propone la aplicación de técnicas de minería de datos para la creación de un modelo de predicción de fallas con el objetivo de aumentar la eficiencia en el proceso de embotellado.

Tabla de cumplimiento de objetivos

Tabla 1.

Cumplimiento de los objetivos del proyecto

OBJETIVO	CUMPLIMIENTO
Identificar a partir del análisis del proceso productivo las características de operación y necesidades y/o requerimientos de mantenimiento.	Capítulo 4
Realizar una revisión de literatura sobre técnicas de minería de datos y sus alcances.	Capítulo 5
Desarrollar un modelo de predicción de fallas para la toma de acciones preventivas a partir de los datos encontrados en el área de embotellado.	Capítulo 7
Generar un modelo estándar para la identificación de fallas en industrias de procesos similares.	Capítulo 8
Elaborar artículo publicable sobre el tema investigado.	Apéndice E

1. Planteamiento del problema

Dada la adquisición de Bavaria S.A por la multinacional ABInBev, se ha generado un cambio de mentalidad en todos los sectores de la empresa enfocándolo especialmente a aspectos relacionados con seguridad, calidad y producción. Esta compañía se identifica con la frase “*primero la calidad, pero es más importante la seguridad*” y ha querido difundirla a todos mediante jornadas de sensibilización, jornadas de integración y mediante capacitaciones, donde se concientiza en lo importante que es la seguridad utilizando un ítem del credo de seguridad: “*llegar a mi casa tan sano como salí*”.

En el área de embotellado la seguridad presenta un riesgo alto ya que durante el proceso productivo se interactúa con máquinas que pueden generar fallas y ocasionar accidentes de trabajo. Esto evidencia la necesidad de mitigar la posibilidad de falla a través de adecuadas ventanas de mantenimiento. Para dicho objetivo, se presenta el método innovador de utilizar técnicas de minería de datos para predecir los posibles daños disminuyendo la probabilidad de un accidente.

Para la aplicación de técnicas de minería de datos, la gerencia del área de embotellado de Bavaria Bucaramanga mantiene un registro histórico de las fallas ocurridas durante la operación. Estas fallas son reportadas por los líderes de equipo al finalizar cada turno de producción en el formato llamado *Informe de turno*. Este registro se lleva mediante el software SAP y es utilizado exclusivamente para la obtención de indicadores del área, especialmente el GLY¹ y LEF², los

¹ Gross Line Yield

² Line Efficiency

cuales están ligados al rendimiento de las líneas de embotellado; y para realizar análisis basados en consultas simples a tal repositorio, esto hace que el potencial de la información almacenada no sea explotado como debería, lo que ocasiona que se ignoren las relaciones existentes entre las variables almacenadas.

Los valores objetivos globales de los indicadores GLY y LEF son 75% y 95% respectivamente, pero actualmente se encuentran por debajo de dichos valores, específicamente en 69% y 93% respectivamente. Por ende, se evidencian oportunidades de mejora en diferentes campos de acción, como en este trabajo que se analizan los tiempos perdidos por fallas en máquinas. Adicionalmente, la capacidad instalada actual genera en promedio 100.000 botellas/hora teniendo como capacidad máxima instalada de 144.000 botellas/hora, es decir, se evidencia una oportunidad de mejora buscando disminuir la brecha entre la capacidad máxima instalada y la capacidad actual.

Buscando obtener mayor conocimiento de la naturaleza de las fallas y las relaciones entre las variables almacenadas, la gerencia del área de embotellado desea aplicar técnicas de minería de datos para descubrir patrones y estructuras ocultas entre dichas variables con el fin de desarrollar un modelo que permita predecir fallas para optimizar la gestión en las ventanas de mantenimiento de las líneas de embotellado y aumentar la confiabilidad en la operación.

2. Justificación

La implementación de técnicas de minería de datos en procesos industriales permite dar apoyo para la elaboración de planes de mantenimiento capaces de brindar funcionamiento continuo en el

proceso. El diseño adecuado de ventanas de mantenimiento genera optimización de los recursos de la empresa, focalización de esfuerzos en áreas críticas que pueden presentar fallas en el siguiente ciclo de producción, aumentar la seguridad de los trabajadores y brindar productos de excelente calidad.

El GLY y LEF son indicadores importantes en empresas de procesos continuos como es el caso del salón de embotellado de Bavaria S.A. Entre los objetivos principales del área se encuentra la obtención de estos indicadores en los niveles más elevados posibles, para lograrlo, es importante detectar la forma como se relacionan las variables registradas en el programa SAP de los turnos de producción, ya que de allí se pueden identificar patrones y tomar acciones preventivas sobre los equipos de mayor vulnerabilidad.

Dicho lo anterior, la aplicación de minería de datos en los valores registrados permite una mayor interpretación de las variables estudiadas incorporando conocimiento importante para la toma de decisiones por parte de la gerencia en la asignación de recursos para garantizar el funcionamiento del proceso aumentando la rentabilidad de la operación.

Para la universidad, la aplicación de áreas académicas en la industria ayuda a fortalecer el lazo de cooperación para futuros estudios. De acuerdo con la revisión de literatura, no se evidencian investigaciones a fin con el tema investigado, por ende, este trabajo será pionero en la aplicación de minería de datos en el proceso de envasado de bebidas brindando ventaja competitiva para la compañía.

3. Objetivos

3.1 Objetivo general

Aplicar técnicas de minería de datos para la identificación de un modelo de predicción de fallas en líneas de embotellado de la industria cervecera “Caso Bavaria Bucaramanga”.

3.2 Objetivos específicos

- ✓ Identificar a partir del análisis del proceso productivo las características de operación y necesidades y/o requerimientos de mantenimiento.
- ✓ Realizar una revisión de literatura sobre técnicas de minería de datos y sus alcances.
- ✓ Desarrollar un modelo de predicción de fallas para la toma de acciones preventivas a partir de los datos encontrados en el área de embotellado.
- ✓ Generar un modelo estándar para la identificación de fallas en industrias de procesos similares.
- ✓ Elaborar artículo publicable sobre el tema investigado.

4. Caracterización del sistema productivo

4.1 Bavaria S.A

La empresa Bavaria S.A es la principal compañía de bebidas del país y una de las marcas con mayor tradición en Colombia. Con más de 128 años de historia, Bavaria sobresale por la calidad de sus productos y el aporte al deporte, a la preservación del medio ambiente, al progreso de las comunidades y al emprendimiento colombiano (Bavaria S.A, 2016).

Bavaria forma parte de la familia ABInBev desde el 2016. Cuenta con una capacidad de producción de aproximadamente 26 millones de Hl/año³. Las marcas producidas son: Águila, Águila Light, Águila Cero, Bahía, Club Colombia, Costeña, Costeñita, Pilsen, Póker, Redd's, Cola & Pola y Pony Malta.



Figura 1. Logotipo Bavaria y ABInBev. Adaptado de Área de embotellado Bavaria Bucaramanga.

Bavaria S.A después de ser adquirida por ABInBev, adapta un credo de seguridad así: **Porque me quiero y amo a mi familia, hoy me comprometo a:** (1) Cuidarme y cuidar a mis compañeros, (2) Asumir comportamientos seguros, utilizando los elementos de protección personal, (3) Velar porque mi área de trabajo siempre sea segura, (4) Reportar los actos y condiciones inseguras y (5)

³ Hectolitros al año

Llegar a mi casa tan sano como salí. Este credo es parte de la nueva cultura que desea implementar esta compañía para disminuir el número de accidentes de trabajo y, por ende, el aumento de la productividad.

4.2 Cervecería de Bucaramanga

Actualmente Bavaria S.A cuenta con seis plantas cerveceras⁴, dos malterías⁵, dos fábricas de etiquetas y una de tapas, distribuidas estratégicamente en todo el país.

Una de las plantas cerveceras se encuentra ubicada en la Ciudad de Bucaramanga, conocida como “La Cervecería Bonita”, con dirección Km 4 carretera Café Madrid e inaugurada el 1° de diciembre de 1948. Cuenta con una capacidad anual de producción de 2.3 millones de Hl / año.

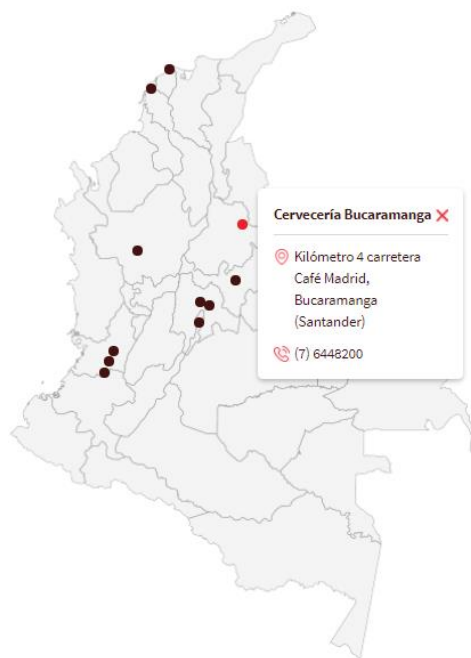


Figura 2. Ubicación geográfica Cervecería Bucaramanga. Adaptada de la página web <http://www.bavaria.co/acerca-de-nosotros/cerveceria-bavaria-bucaramanga>

⁴ Barranquilla, Bucaramanga, Duitama, Medellín, Tocancipá y Yumbo.

⁵ Cartagena y Tibitó.

El área de embotellado de la cervecería de Bucaramanga está conformada por dos (2) líneas de producción, línea 1 y línea 2. Para este proyecto, se analizan los datos que corresponden a la línea 2 porque presentan mayor número de horas en operación, se elaboran productos con igual volumen (330 ml) y las máquinas son más nuevas con relación a la línea 1. La línea de interés elabora los siguientes productos: pony malta, light, poker, pilsen, águila y club colombia.

4.3 Área de Embotellado de Bavaria Bucaramanga



Figura 3. Instalaciones área de embotellado. Adaptado de área de embotellado Bucaramanga

4.3.1 Estructura organizacional. El orden jerárquico del equipo de trabajo se evidencia en la Figura 4. Está conformado por 14 personas, cinco (5) profesionales en el área de mantenimiento y siete (7) en el área de producción, contando su respectivo coordinador. De los líderes de turno de producción, a cada línea pertenecen tres.

4.3.2 Descripción del proceso de envasado. El proceso inicia en la Depaletizadora, a continuación, viene la etapa de Desempacado, sigue la etapa del lavado de las botellas, estas continúan a los inspectores de envase vacío y luego al paso de envasado y tapado en las envasadoras, continua con su respectivo fechado para seguir con las etapas de pasteurización y etiquetado. Las botellas llenas se inspeccionan, por etiquetado, nivel y tapado.

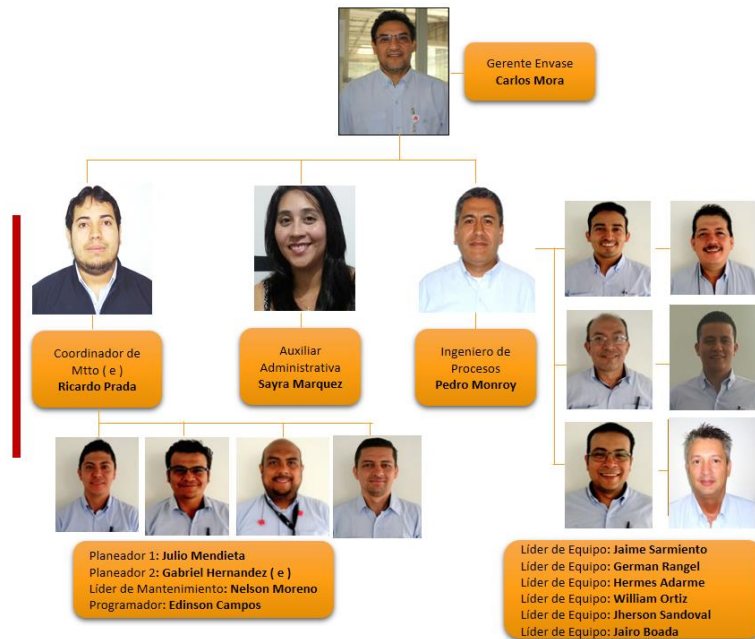


Figura 4. Estructura organizacional equipo embotellado. Adaptado del Área de Embotellado de Bavaria Bucaramanga



Figura 5. Descripción del proceso de embotellado. Adaptado del Área de Embotellado de Bavaria Bucaramanga

Se ponen las botellas en las cajas mediante el proceso de empaclado, seguidamente se verifica la cantidad exacta en cada caja para finalizar con el proceso de paletización, cerrando así el ciclo completo del embotellado de la cerveza. Existe un subproceso que forma parte del proceso que es el lavado de las cajas correspondiente a la máquina lavadora de cajas. La descripción gráfica del proceso se encuentra en la figura 5.

Cada paso del ciclo de embotellado consta de equipos encargados de la ejecución de su respectiva fase. A continuación, cada uno de estos equipos serán ilustrados y justificando la función que realizan en el proceso.

Depaletizadora: Una vez recibida las estibas⁶, esta máquina (ver figura 6) es encargada de remover automáticamente capas⁷ de cajas con botellas vacías y enviarlas a la desempacliadora por medio de un transportador. Las estibas vacías son enviadas al estibador para su posterior uso en la Paletizadora.



Figura 6. Depaletizadora. Adaptado del Área de Embotellado de Bavaria Bucaramanga

⁶ Cada estiba contiene cinco (5) capas para envase 330 ml.

⁷ Cada capa consta de nueve (9) cajas organizadas de 3x3.

Desempacadora: Esta máquina tiene como objetivo remover las botellas vacías de las cajas y alimentar el transportador a la lavadora. Las cajas vacías son enviadas a la lavadora de cajas para su posterior subproceso.



Figura 7. Desempacadora. Adaptado del Área de Embotellado de Bavaria Bucaramanga

Lavadora de Botellas: Esta máquina es una de las más importantes del proceso de embotellado. Su objetivo es lavar las botellas retornables usadas o nuevas, remover etiqueta y enjuagar las botellas. Las botellas lavadas son enviadas al inspector de vacías mediante un transportador de botellas.



Figura 8. Lavadora de botellas. Adaptado del Área de Embotellado de Bavaria Bucaramanga

Inspector de Botellas Vacías: Este equipo electrónico (figura 9) tiene el objetivo de hacer la inspección de las once zonas previamente establecidas a cada una de las botellas y establecer si es apta o no para ser llenadas. Adicional a la parte electrónica, cuenta con un sistema mecánico que dirige el envase bien lavado hacia la envasadora y el envase fuera de especificaciones hacía una mesa de rechazo.



Figura 9. Inspector de botellas vacías. Adaptado del Área de Embotellado de Bavaria Bucaramanga

Envasadora: Máquinas cuyo objetivo es llenar las botellas con cerveza manteniendo la calidad y especificaciones del producto manteniendo niveles correctos y repetitivos minimizando la pérdida de cerveza. En la figura 10 se muestra el acople entre la envasadora y el coronador⁸. Seguido a esto, se encuentra el fechador encargado de colocar la fecha de elaboración del producto.

⁸ Sección acoplada a la envasadora encargada de colocar la tapa en la botella.

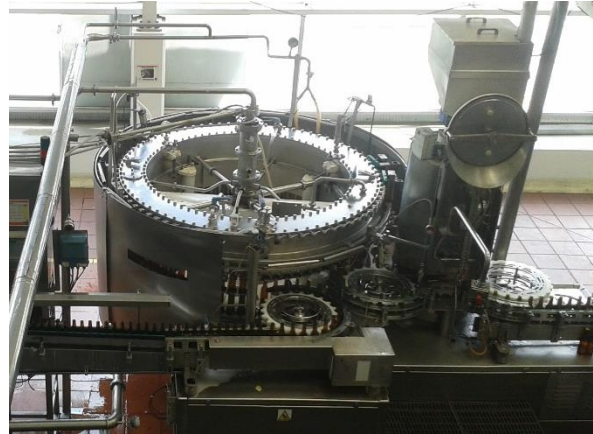


Figura 10. Envasadora y coronador. Adaptado del Área de Embotellado de Bavaria Bucaramanga

Pasteurizadora: El principal objetivo es eliminar y/o reducir a niveles controlables microorganismos que puedan ser riesgosos para la salud humana o para la calidad del producto. El envase es gradualmente calentando hasta llegar a la temperatura letal, es mantenida por algunos minutos y luego enfriada para terminar con una temperatura en la cual la cerveza no sufra deterioro.



Figura 11. Pasteurizador. Adaptado del Área de Embotellado de Bavaria Bucaramanga

Etiquetadora: Máquina giratoria cuyo objetivo principal es adherir pegante a la etiqueta y colocarla a las botellas, brindando excelente presentación al envase.



Figura 12. Etiquetadora. Adaptado del Área de Embotellado de Bavaria Bucaramanga

Inspector de producto terminado: Máquina encargada de verificar la altura del llenado, tapado y la cantidad exacta de etiquetas en cada botella. En caso de estar fuera de especificaciones en alguna de las tres variables, esta botella es rechazada de forma automática por la máquina. Cuando el defecto es por ausencia de etiqueta, estos productos son reprocesados de forma manual.



Figura 13. Inspector de producto terminado. Adaptado del Área de Embotellado de Bavaria Bucaramanga

Lavadora de cajas: El objetivo principal de esta máquina es recibir las cajas sucias y mediante cuatro pasos de lavado a diferentes temperaturas remover objetos extraños, basura y lavar por dentro y por fuera antes de ser usadas nuevamente.



Figura 14. Lavadora de cajas. Adaptado del Área de Embotellado de Bavaria Bucaramanga

Empacadora: Máquina con función inversa a la desempacadora. Su objetivo es colocar las botellas de cerveza terminada y etiquetada en las cajas plásticas.



Figura 15. Empacadora. Adaptado del Área de Embotellado de Bavaria Bucaramanga

Inspector de caja llena: El objetivo de este equipo es inspeccionar cada caja llena proveniente de la empacadora. Si falta alguna botella, será rechazada de forma automática para completarla de forma manual.



Figura 16. Inspector de caja llena. Adaptado del Área de Embotellado de Bavaria Bucaramanga

Paletizadora: Esta máquina es igual a la depaletizadora, pero con la función inversa. Se encarga de recibir y ordenar las cajas llenas, para luego transportarlas y agruparlas en las estibas. Estas estibas son llevadas por los montacargas al lugar de almacenamiento.



Figura 17. Paletizadora. Adaptado del Área de Embotellado de Bavaria Bucaramanga

4.3.3 Rendimiento de una línea. La figura 18 permite observar el rendimiento esperado por cada máquina para obtener un rendimiento efectivo de la línea. Este gráfico permite identificar cuáles son las máquinas de mayor interés, permite una visión general de toda la línea: velocidades, interacción con otras máquinas y confiabilidad. Adicionalmente, es una herramienta clave para observar deficiencias en la línea.

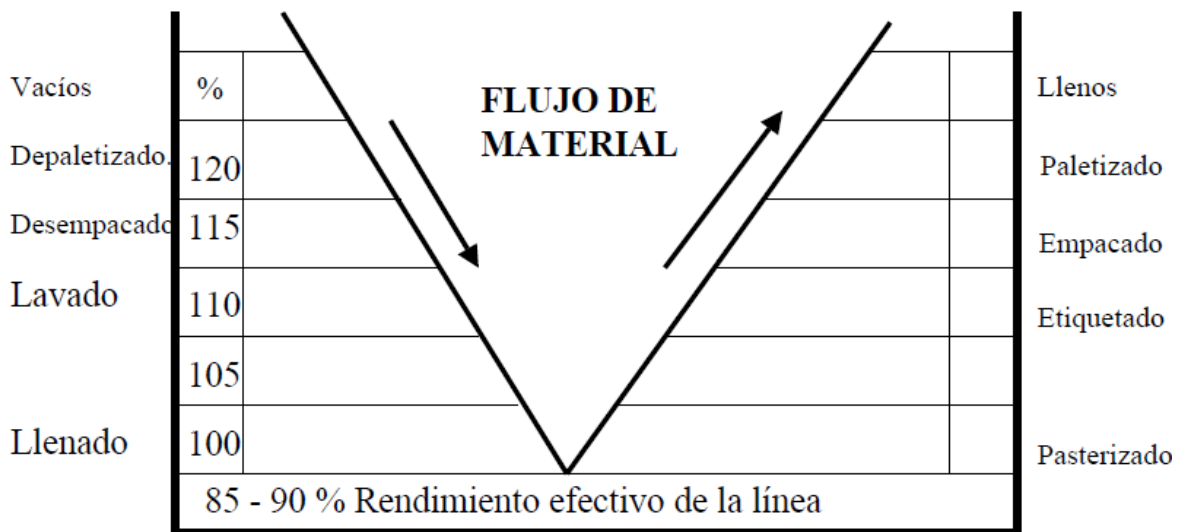


Figura 18. Gráfico en V. Adaptado del Área de Embotellado de Bavaria Bucaramanga

Los requerimientos de mantenimiento se basan en la metodología RCM (Mantenimiento Centrado en la Confiabilidad) definido como una disciplina lógica enfocada al proceso de falla para desarrollar un programa de mantenimiento programado que determinará los niveles inherentes de seguridad y confiabilidad de equipos complejos a costos mínimos.

Estos requerimientos tienen en cuenta la siguiente secuencia de pasos para cada equipo:

1. Falla de los elementos de la planta.
2. Definir la función estándar de desempeño asociado a cada elemento.

3. Identificar posibles fallas funcionales para cada elemento.
4. Establecer las causas “más probables” para todas las fallas funcionales.
5. Establecer el efecto y la secuencia de cada modo de falla.
6. Seleccionar la política de manejo de fallas más apropiada para cada modo de falla.
7. Desarrollar cronograma con base en el paso 6.

5. Marco teórico

En esta sección se expondrán términos de gran importancia para el desarrollo del modelo de predicción de fallas utilizando técnicas de minería de datos utilizadas para pronosticar el tiempo perdido en un ciclo de producción. Para ello, se comienza con un análisis preliminar de literatura, seguido de la definición de algunos conceptos básicos, se comentan las formas de mantenimiento que maneja el área de embotellado. Luego se hace referencia a todo el proceso de Descubrimiento de Conocimiento de Bases de Datos. Finalmente, se nombrarán las dos técnicas utilizadas en este proyecto: Análisis de Componentes Principales y Regresión Lineal Múltiple.

5.1 Análisis preliminar de literatura

Son muchos los motivos que llevan a trabajar con la información de los procesos, ya que dándole un correcto uso se pueden utilizar para controlar, planificar, investigar, predecir y tomar decisiones de acuerdo con el área de interés. Además, como los datos recopilados en el transcurso del tiempo aumentan la base de datos, los métodos de análisis estadísticos tradicionales pueden ser

de poca ayuda para descubrir los patrones escondidos en las bases de datos y la necesidad de un nuevo enfoque parece inevitable (Harding, 2001)

El proceso de Descubrimiento de Conocimiento en Bases de Datos se aplica en la actualidad debido a la madurez de los tres elementos que lo conforma: herramientas que permiten la colección masiva de datos, computadores potentes con multiprocesadores y algoritmos para aplicar el modelo (Vélez Torres & Hernández R, 2001). En general, se han implementado diversas aplicaciones en diferentes campos de acción: Ingeniería, telecomunicaciones, medicina, fuga de cliente, clasificación de consumidores, controles de producción, pronósticos de ventas y, por último, para predicción de fallas.

Según (Mosaddar & Shojaie, 2013) el entorno empresarial competitivo lleva a la compañías a reducir los costos a fin de crear ventaja competitiva y ampliarla, es por eso que, diferentes áreas de los negocios han sido excavadas para un método mejor y eficiente para realizar las tareas y por lo tanto reducir los costos de todo el sistema.

De acuerdo a (Choudhary, Harding, & Tiwari, 2009) la aparición de métodos de mantenimiento preventivo y predictivo o Total Productive Maintenance (TPM) ha sido enfocado a la necesidad de mejorar la calidad y por lo tanto, la efectividad de las actividades de mantenimiento deben ser realizadas de la mejor manera posible para disminuir falla en el equipo y sus costos asociados.

(Han, Kamber, & Pei, 2012) clasifican las técnicas de minería de datos en dos categorías: descriptivo y predictivo. Las tareas descriptivas pueden encontrar patrones ocultos, tendencias y correlaciones entre variables de datos, mientras que el predictivo usan datos existentes para predicciones futuras (Larose, 2004).

El trabajo realizado por (Gómez Torrejano, 2009) nos presenta la aplicación de la minería de datos para proponer un modelo predictivo de ocurrencias de fallas en los equipos y sistemas de la plataforma tecnológica de PDVSA AIT Servicios Comunes Centro. Abordando otra tipo de industria, (Kang et al., 2017) presenta en análisis de las relaciones causales entre el producto final y los reclamos de los clientes que adquieren dicho producto, con el fin de identificar las fallas comunes y evitar que productos posteriores tengan el mismo defecto presentado en el reclamo.

De acuerdo con la revisión bibliográfica realizada, en ninguno de los 178 artículos identificados con la ecuación de búsqueda se evidencian aplicaciones a la industria cervecera, específicamente a las líneas de embotellado. Por el contrario, se encuentran trabajos relacionados a industrias como telecomunicaciones, energía eólica, petroleras, servicio al cliente, redes inalámbricas, construcción de vías, académicas y ventas. Por tal razón, es una excelente oportunidad para observar el comportamiento en este sector y convertirla en una ventaja competitiva.

5.2 Conceptos básicos

A continuación, se definen algunos conceptos iniciales de acuerdo con el diccionario de la Real Académica Española DRAE:

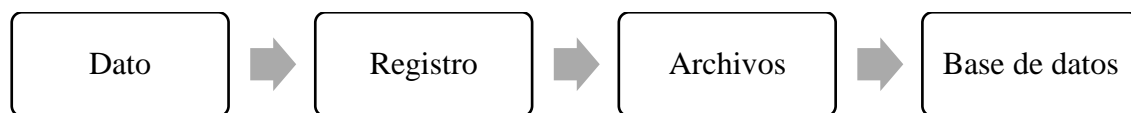


Figura 19. Relación datos, registro, archivo y bases de datos

Dato: Información sobre algo concreto que permite su conocimiento exacto o sirve para deducir las consecuencias derivadas de un hecho o detalle que sirve de base a un razonamiento o una investigación. Asimismo, se puede decir que el dato es una representación simbólica, atributo o

característica de una entidad. El dato no tiene valor semántico (sentido) en sí mismo, pero convenientemente procesado se puede utilizar de cálculo o toma de decisiones.

Registro: Libro, a manera de índice, donde se apuntan datos, de tal forma que cada uno de sus componentes ocupa un espacio denominado campo. En esta investigación, los registros se refieren a la información almacenada por los equipos de producción para conservar los reportes de fallas y sus respectivas medidas de acción tomadas.

Archivos: Conjunto ordenado de documentos que una persona, una sociedad, una institución, etc., producen en el ejercicio de sus funciones o actividades. Para este estudio, los archivos son los registros de los informes de turno e informes de ventanas de mantenimiento.

Base de datos: Conjunto de datos organizados de tal modo que permita obtener con rapidez diversos tipos de información. Recolección organizada en forma cronológica de archivos que permita su utilización de forma sencilla y rápida. En este caso, la base de datos será utilizada para desarrollar un modelo que permita predecir de forma sencilla las posibles fallas que se ocurrirán en la siguiente semana de producción.

Adicional a la definición dada para cada término, hay que entender la relación existente entre estas palabras, comenzando que para obtener una generosa base de datos es indispensable la excelente selección de datos para sus respectivos tratamientos. En la Figura 19 nos muestra el orden adecuado en la creación de una base de datos, es decir, los registros son un conjunto de datos, los archivos una colección de registro y la base de datos un conglomerado de archivos.

Al tener un conglomerado de las fallas que han ocurrido en un lapso determinado, se puede tener un control centralizado de los datos y mediante la utilización de modelos se pueden establecer patrones no vistos directamente y que si pueden ser reconocidos mediante el modelo. Igualmente,

la creación de variables que puedan dar información adicional de la relación entre estas y de gran interés para resolver problemas encontrados en el proceso.

5.3 Formas de mantenimiento

La confiabilidad inherente en los equipos es el nivel de confiabilidad que se logra cuando existe un programa de mantenimiento efectivo. En el área existen cuatro (4) formas de mantenimiento:

Reemplazo en vez de mantenimiento: Esta política aplica a los equipos pequeños, económicos y de fácil reemplazo – no diseñados para durar, se deben botar cuando fallan.

Reemplazo planeado: Esta política aplica a las unidades individuales, como por ejemplo herramientas de máquinas, compresores móviles y automóviles. Se compran equipos de buena calidad y sólo se realiza mantenimiento simple. Al final, se vende, chatarra o alquila el elemento, antes de que pierda su eficiencia operativa

Mantenimiento por falla: Esta política aplica a todos esos elementos que no afectan el proceso de producción, o el costo en lo que realizar el mantenimiento preventivo es mayor al costo de la pérdida de producción, teniendo en cuenta todos los factores. Este tipo de mantenimiento se realiza únicamente después de la falla, a excepción de la lubricación y de un mantenimiento sencillo.

Mantenimiento preventivo: Aplica a cualquier proceso que sea continuo o altamente automatizado. La base de esta política es evitar fallas.

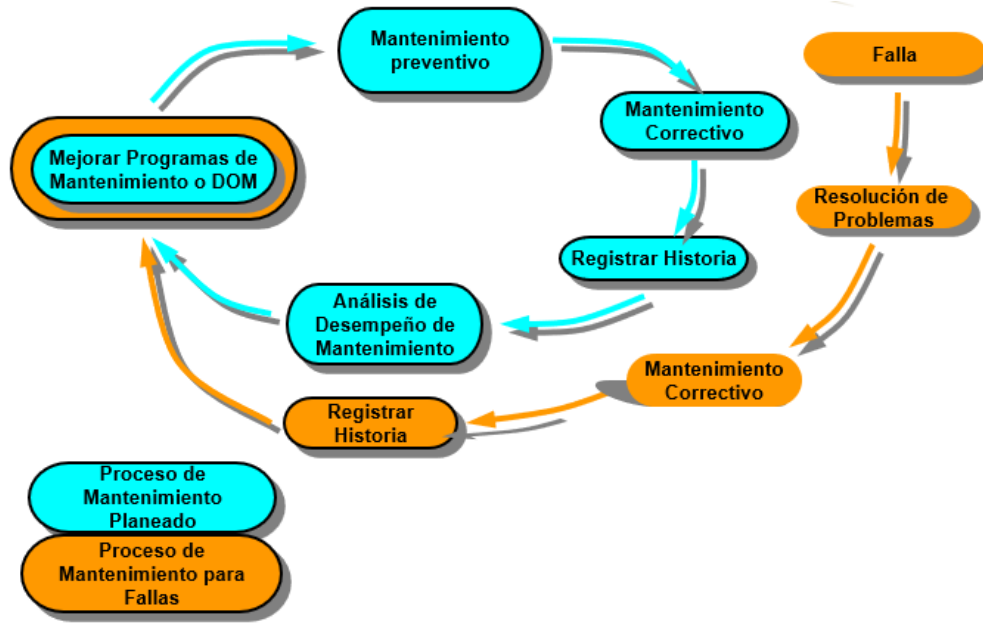


Figura 20. Diagrama de mantenimiento. Adaptado del Área de Embotellado de Bavaria Bucaramanga

En la figura 20 se observa que el mantenimiento preventivo es esa fase que toda industria de procesos continuos desea implementar en sus líneas de producción. Adicional a esto, no se debe entender el mantenimiento como solo hacer intervenciones para la conservación del bien, sino que se tiene que involucrar la participación en la mejora continua de los procesos productivos, acompañados de la evolución tecnológica

Algunas de las ventajas de un mantenimiento predictivo son que el equipo se encuentra en funcionamiento mientras se monitorea y permite que todos sus elementos cumplan su vida útil. De la misma forma, un buen mantenimiento predictivo se justifica cuando (1) se logra reducir el riesgo de falla múltiple a un nivel aceptable, (2) cuando se da un tiempo suficiente de advertencia que permita actuar al equipo, (3) en las consecuencias operacionales si el costo de monitoreo es menor que el costo de no hacerlo y (4) en consecuencias no operacionales si el costo de monitoreo más el costo de reparar es menor que el costo de sólo reparar (Bavaria S.A, 2014).

5.4 Descubrimiento de conocimiento en bases de datos

Esta metodología permite, a partir del uso de técnicas de minería de datos, obtener información oculta mediante la interacción con grandes volúmenes de datos y utilizados para mejorar los procesos productivos de una empresa. Este proceso es definido por (Fayyad, Piatetsky-Shapiro, & Smyth, 1996) como “proceso no trivial de identificar patrones en una forma valiosa, novedosa, potencialmente útiles y comprensibles a partir de los datos”.

Existen seis elementos esenciales que califican el descubrimiento del conocimiento como una sola técnica. Las siguientes son las características básicas que comparten dichos enfoques (Fayyad et al., 1996):

- Todos los enfoques tratan con grandes cantidades de datos.
- Se requiere eficiencia en el método empleado en el descubrimiento de conocimiento, debido al volumen de datos.
- La exactitud de los resultados que se obtienen es un elemento esencial.
- Todos requieren el uso de un lenguaje de alto nivel.
- Todos los enfoques usan alguna forma de aprendizaje automático.
- Todos producen algunos resultados interesantes.

El proceso que se muestra en la Figura 21, involucra los pasos más importantes que interactúan con el usuario para el proceso de toma de decisiones.

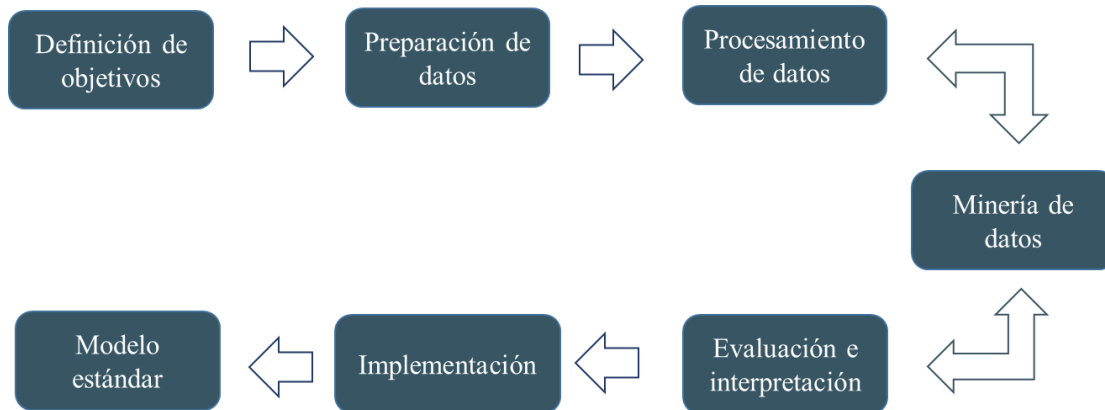


Figura 21. Proceso de Descubrimiento de Conocimiento en Bases de datos. Adaptado de la metodología CRIP-DM y KDD



Figura 22. Metodología SMART. Adaptado de área de embotellado Bucaramanga

Las etapas de formar parte del proceso se describen a continuación:

Definición de objetivos: la metodología SMART es considerada desde la parte productiva como la de mayor eficacia porque evita incertidumbres y ofrece toda la información con claridad para que los objetivos sean cumplidos con éxito.

La figura 22 presenta la definición de cada una de las variables, donde **S** hace referencia a que se debe expresar con claridad cuáles son los resultados que se quieren conseguir; **M** se refiere a que se deben establecer indicadores para las propuestas de mejora donde se pueda hacer un

seguimiento en el tiempo, identificando si las propuestas fueron exitosas, fracasaron o no generaron ningún cambio; **A** aclarece que los objetivos si se van a poder cumplir, hay que tener en cuenta los costos derivados, el tiempo y el esfuerzo necesario para la consecución de los objetivos; **R** hace referencia a que lo objetivos planteados deben ser relevantes y estar direccionados con el plan de la empresa y **T** hace énfasis en que los objetivos deben tener una fecha límite para que el equipo de trabajo pueda establecer un cronograma de actividades para alcanzar dichos objetivos.

Preparación de datos: Según (Dorian Pyle, 1999) “la minería de datos utiliza principalmente fuentes de datos disponibles como data warehouses, data mart, bases de datos o archivos de conglomerado de datos”. Consiste en seleccionar y obtener el conjunto de datos con los que se van a trabajar. De acuerdo a (Vélez Torres & Hernández R, 2001) “los tipos de datos seleccionados pueden ser organizados en tablas bajo ciertas condiciones y para ciertas operaciones de la minería de datos, tales como la creación de un modelo de clasificación, o de regresión, se puede utilizar solamente una muestra de datos, ya que un muestreo resulta menos costoso que un censo”.

Procesamiento de datos: Los datos en la realidad son sucios, incompletos e incoherentes. Muchas veces existen datos que no son de interés para una investigación y en otras ocasiones se necesita la creación de nuevas variables para resolver el problema planteado. El objetivo principal de este paso es obtener los datos finales para la investigación y que estos sean de calidad y generen valor a la investigación. De acuerdo con (Paolo, 2003) existen estas técnicas básicas para el procesamiento de datos:

- Limpieza de datos: implica técnicas para completar los datos perdidos, suavizar el ruido, manejar valores atípicos, detectar y eliminar datos redundantes.

- Transformación de datos: técnicas que pone los datos en forma apropiada para su respectivo manejo.
- Reducción de datos: Se utiliza para reducir el conjunto de datos que se extrajo. Mientras técnicas de “reducción de dimensiones” eliminar atributos innecesarios, las técnicas de “comprensión de datos” y “reducción de numerosidad” proporcionan otras formas de representación de datos reducidos.
- Discretización: es una forma de reducción de datos, reduce el número de niveles de un atributo mediante la recopilación y sustitución de conceptos de bajo nivel por conceptos de alto nivel.

Minería de datos: Una vez identificado lo que se desea obtener, se debe proceder a decidir cuáles son las técnicas más apropiadas para desarrollar el modelo de predicción de fallas. Cabe resaltar que durante la utilización de las técnicas puede resultar necesario volver al paso anterior para la creación o eliminación de variables.

Evaluación e interpretación: Desarrollado el modelo predictivo se debe corroborar que efectivamente funciona y evaluarlo mediante el modelo propuesto. Así mismo, se debe interpretar de forma clara qué significa cada resultado arrojado por la ecuación. Estos resultados deben ser lo más claros y sencillos posibles para la fácil interpretación de las personas que lo van a utilizar en la empresa y que no tienen conocimiento sobre la esencia del modelo.

Implementación: En este paso, las personas encargadas en la empresa de las áreas de mantenimiento y producción deben poner en funcionamiento dicho modelo mediante los datos de la próxima semana de producción para verificar las posibles fallas que se presentarán. Una vez

terminada la semana se hace un símil entre lo predictivo y lo real, si este concuerda se considera como correcto y se correrá nuevamente el modelo para la próxima semana y de acuerdo con los resultados se tomarán medidas preventivas en la ventana de mantenimiento anterior a dicha semana de producción.

Modelo estándar: Dado que Bavaria S.A es una empresa de productos de consumo masivo, así mismo, hay muchas empresas en el país que poseen procesos similares y que se pueden adaptar a la metodología usada en este trabajo. Entonces, este paso consiste en estandarizar el modelo de forma tal que pueda ser apropiado por una empresa y adaptarlo de acuerdo con el proceso que manejen.

5.5 Análisis de Componentes Principales

Al momento de iniciar un análisis, se dispone de mucha información donde están presentes muchas variables, por ejemplo, si se tienen 20 variables se deben de considerar alrededor de 180 coeficientes de correlación. Es evidente que, si se toman demasiadas variables es normal que estén relacionadas o representen lo mismo entre diferentes puntos de vista. Es por lo que se hace necesario buscar la forma de reducir el número de variables a unas nuevas que representen de forma lineal las variables iniciales.

Dado este inconveniente, aparece la técnica de Análisis de Componente Principales la cual ayuda a dar una solución. Estas nuevas variables son combinaciones lineales de las variables originales y se van construyendo de acuerdo con la variabilidad que presenta la muestra.

Para el cálculo de las componentes se puede hacer en diferentes programas estadísticos, pero a continuación presentaremos la forma manual (Eduardo, 1999).

Se consideran las variables originales $X_1, X_2, X_3 \dots X_i$ y el objetivo es calcular un nuevo grupo de variables $Y_1, Y_2, Y_3 \dots Y_i$ las cuales deben ser incorreladas entre si y cuya varianza decrezca de forma progresiva.

Cada Y_m ($m = 1, 2, \dots, i$) es una combinación lineal de las $X_1, X_2, X_3 \dots X_i$ iniciales

$$Y_m = b_{m1}x_1, b_{m2}x_2, b_{m3}x_3 \dots b_{mi}x_i = b_m^o x$$

Donde $b_m^o x = (b_{1m}, b_{2m}, b_{3m} \dots b_{im})$ es un vector de constantes, y $x = \begin{pmatrix} x_1 \\ \dots \\ x_i \end{pmatrix}$

El siguiente paso es maximizar la varianza. Para ello se debe mantener la ortogonalidad de la transformación haciendo que el módulo del vector $b_m^o x = (b_{1m}, b_{2m}, b_{3m} \dots b_{im})$ sea 1. Es decir,

$$b_m^o \cdot b_m = \sum_{k=1}^i b_{km}^2 = 1$$

Entonces, el primer componente se calcula eligiendo b_1 de modo que Y_1 tenga la mayor varianza, sujeta a la restricción $b_m^o \cdot b_m = 1$. Para el segundo componente principal se calcula obteniendo b_2 de modo que la variable obtenida Y_2 este incorrelada con Y_1 .

Habitualmente se conservan los componentes que recogen la mayor parte de la variabilidad (varianza mayor a 1). En muchas ocasiones estas variables permiten la representación en dos o tres dimensiones con el fin de identificar grupos naturales entre las observaciones.

Los objetivos más importantes de todo análisis por componentes principales son:

- Generar nuevas variables que puedan expresar la información contenida en el conjunto original de datos.

- Reducir la dimensionalidad del problema que se está estudiando, como paso previo para futuros análisis.
- Eliminar, cuando sea posible, algunas de las variables originales si ellas aportan poca información.

Las nuevas variables generadas se denominan *componentes principales* y poseen algunas características estadísticas deseables, tales como independencia (cuando se asume multinormalidad) y en todos los casos no correlación. Esto significa que, si las variables originales no están correlacionadas, el análisis por componentes principales no ofrece ventaja alguna (Pla, 1986).

5.6 Regresión Lineal Múltiple

A diferencia de la regresión lineal simple, esta permite utilizar más de una variable explicativa, por ende, las estimaciones son más precisas (Ortiz, 2015).

De acuerdo con (Barón López & Télez Montiel, 2004) la aplicación real de la aplicación de este técnica en las investigaciones son:

- Identificación de variables explicativas. Ayuda a crear un modelo donde se seleccionan las variables que puedan influir en la respuesta, descartando aquellas poco significativas.
- Detección de interacciones entre variables independientes que afectan a la variable respuesta.
- Identificación de variables confusoras. Es un problema difícil el de su detección, pero de interés en investigación no experimental, ya que el investigador no tiene control sobre las variables independientes.

De acuerdo con (Barón López & Télez Montiel, 2004) hay que tener en cuenta ciertos requerimientos para poder usar esta técnicas: linealidad, normalidad y distribución de los residuos, número de variables dependientes, colinealidad y observaciones anómalas.

De acuerdo con (García & López, 2012) el modelo de regresión múltiple para n variables predictoras sería como el que se muestra a continuación

$$y = b_0 + b_1x_1 + b_2x_2 \dots + b_nx_n$$

Para encontrar los b_i se plantea el modelo en términos de matrices como se muestra a continuación

$$Z = \begin{bmatrix} z_{11} & \dots & z_{1n} \\ z_{21} & \dots & z_{2n} \\ \dots & \dots & \dots \\ z_{m1} & \dots & z_{mn} \end{bmatrix}; Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_m \end{pmatrix}; B = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix}$$

En la matriz Z, las filas representan los m valores disponibles para calcular la regresión y las columnas los n atributos que formarán parte de la regresión. EL vector Y está formado por los valores de la variable dependiente para cada uno de los valores, y el vector B es el que se desea calcular, ya que corresponde a los parámetros desconocidos necesarios para construir el modelo de regresión múltiple. Representando con X^T la matriz transpuesta de X y con X^{-1} inversa de la matriz X, se calculará el vector B, así:

$$B = (Z^T Z)^{-1} Z^T Y$$

Para determinar si la recta de regresión lineal múltiple está bien ajustada, se emplea el mismo concepto de la regresión lineal simple, así:

$$R^2 = 1 - \frac{(Y - ZB)^{-1}(Y - ZB)^T}{\sum_{i=1}^m (y_i - \bar{y})^2} \text{ este valor oscila entre 0 y 1}$$

6. Proceso de descubrimiento de conocimiento en bases de datos

El presente capítulo expone el proceso de descubrimiento de conocimiento en bases de datos a través de sus etapas principales, con el objetivo de brindar una visión global de la data para la solución del problema planteado. Inicia con la etapa de definición de objetivos en donde se identifican que resultados son los que se desean obtener con base a las necesidades de la industria; continúa con la etapa de preparación de datos y su respectivo procesamiento para luego finalizar con la etapa de minería de datos donde se presenta el uso de las diferentes técnicas estadísticas que se utilizaron para el cumplimiento de los objetivos propuestos.

6.1 Definición de objetivos

El proceso de descubrimiento de conocimiento en base de datos inicia con la identificación de las necesidades de la empresa para la creación de una metodología que lleve a encontrar una solución idónea al problema. Para el proyecto se plantearon los siguientes objetivos:

- Identificación de máquinas, meses y productos críticos y la relación entre ellos.
- Crear grupos de máquinas relacionadas con base a los tiempos perdidos para desarrollar estrategias de mantenimiento.
- Pronosticar el tiempo total perdido de acuerdo con las diferentes variables manejadas en el sector.

6.2 Preparación de los datos

En esta etapa, se identifican las fuentes donde se extraerán los datos, ya sea de servidores internos o de grandes volúmenes de archivos almacenados por el área. Para este proyecto, se

dispondrá de los datos históricos de producción de los años 2015, 2016 y 2017 del área de embotellado de la empresa Bavaria S.A sede Bucaramanga.

Los datos se encuentran en el programa SAP-PM y son extraídos en archivos tipo “Excel”. Una vez finalizada la extracción de los datos, se recopilaron los siguientes archivos:

- Datos Zenvase2015-2017 línea 2 (ver en CD)
- Datos Ztiempos2015-2017 línea 2 (ver en CD)
- Datos regresión lineal múltiple (ver en CD)

Estos archivos se encuentran en el CD como Anexo B, Anexo C y Anexo D respectivamente.

6.3 Procesamiento de los datos

En esta fase se seleccionan, se limpian y se transforman los datos extraídos con el fin de dejar las variables significativas para el estudio.

6.3.1 Variables eliminadas. Variables eliminadas. Se eliminan las siguientes variables de la base de datos “Datos Zenvase2015-2017 línea 2” porque no son significativas para la solución del problema (ver tabla 2).

6.3.2 Variables modificadas. Las siguientes variables son modificadas de forma que pueda ser trabajada en la etapa de la minería de datos para la extracción de conocimiento (ver tabla 3).

6.3.3 Variables creadas. Con base a los nuevos grupos de productos, se generaron igual cantidad de variables para ser procesadas de forma dicotómica al momento de aplicar regresión lineal múltiple. Igualmente, las unidades totales entregadas en depósitos fueron separadas por producto (ver tabla 4).

Tabla 2.

Variables eliminadas de Zenvase2015-2017 línea 2

VARIABLES ELIMINADAS		
Número de orden	Hectolitros en depósito	Hectolitros envasados
Tiempo programado	Tiempo descontado	Eficiencia total
Tiempo otras causas	Tiempo aseo	Eficiencia mecánica
Tiempo mantenimiento	Horas extras	Eficiencia fábrica

Tabla 3.

Variables modificadas

VARIABLE ANTIGUA	VARIABLE NUEVA
Producto	Grupos de productos
Fecha	Separación de la fecha en tres columnas: semana, mes y año

Tabla 4.

Variables creadas

VARIABLES CREADAS	
Light-P1-Club	Light-P2
Light-P2-Club	Light-P3
Light-P3-Club	Pony-Light-P1-Club
Pony-Light-P2	Pony-Light-P2-Club
Pony-Light-P3	Pony-Light-P3-Club
Unidades Pony	Unidades Light
Unidades Poker	Unidades Pilsen
Unidades Águila	Unidades Club

En la tabla 4 se observa que aparece la letra P acompañado de un número, la letra P hace referencia a que se elabora cerveza de envase marrón (águila, pilsen y poker) y el número se refiere cuantas de estos tipos de cerveza se elaboran independiente del orden. Entonces, se encuentran las siguientes opciones:

- P1 = Se elabora un (1) solo tipo de cerveza de envase marrón
- P2 = Se elaboran dos (2) tipos de cerveza de envase marrón.
- P3= Se elaboran tres (3) tipos de cerveza de envase marrón.

6.3.4 Definición de variables definitivas. A continuación, se definen las variables significativas para el proyecto

- Semana, mes y año: Fecha de elaboración del producto
- Producto: artículo elaborado
- Unidades en depósito: Unidades elaboradas entregadas al depósito
- Depaletizadora: Tiempo perdido por la máquina Depaletizadora
- Desempacadora: Tiempo perdido por la máquina Desempacadora
- Lavadora de botellas: Tiempo perdido por la máquina Lavadora de botellas
- Inspector de botellas vacías: Tiempo perdido por la máquina Inspector de botellas vacías.
- Envasadora: Tiempo perdido por la máquina Envasadora
- Fechador: Tiempo perdido por la máquina Fechador
- Pasteurizador: Tiempo perdido por la máquina Pasteurizador
- Etiquetadora: Tiempo perdido por la máquina Etiquetadora
- Inspector de producto terminado: Tiempo perdido por la máquina Inspector de producto terminado

- Lavadora de cajas: Tiempo perdido por la máquina Lavadora de cajas
- Empacadora: Tiempo perdido por la máquina Empacadora
- Inspector de caja llena: Tiempo perdido por la máquina Inspector de caja llena
- Paletizadora: Tiempo perdido por la máquina Paletizadora
- Transportador de canastas: Tiempo perdido por la máquina Transportador de canastas
- Transportador de botellas: Tiempo perdido por la máquina Transportador de botellas
- Tiempo total: Tiempo total perdido por cada individuo
- Grupos de productos: Conjunto de productos que forman parte de un ciclo de producción.
- Unidades pony: Unidades elaboradas en un ciclo de producción de producto pony
- Unidades light: Unidades elaboradas en un ciclo de producción de producto light
- Unidades poker: Unidades elaboradas en un ciclo de producción de producto poker
- Unidades pilsen: Unidades elaboradas en un ciclo de producción de producto pilsen.
- Unidades águila: Unidades elaboradas en un ciclo de producción de producto águila
- Unidades club: Unidades elaboradas en un ciclo de producción de producto club
- Tiempo total por grupo de productos: Tiempo total perdido por un grupo de productos en un ciclo de producción

6.4 Minería de datos

Mediante esta nueva etapa se pretende generar nuevo conocimiento sobre el tiempo total perdido en una línea de embotellado. Se utilizaron tres técnicas para analizar los datos las cuales están enmarcadas dentro de las necesidades para cumplir los objetivos planteados. La primera se basa en un análisis previo de los datos, la segunda en un Análisis de Componentes Principales y por último Regresión Lineal Múltiple.

Para esto, se utilizaron diferentes herramientas tales como los programas IBP SPSS Statistics y RStudio; los cuales garantizan un rápido procesamiento de los datos y la posibilidad de tomar decisiones en base a los resultados.

6.4.1 Análisis previo de los datos. Para este análisis se utiliza el programa IBM SPSS donde se siguen los pasos descritos a continuación. Una vez cargado los datos en formato de Excel, se sigue la siguiente ruta: analizar > Comparar medias > Anova de un factor (ver figura 23).

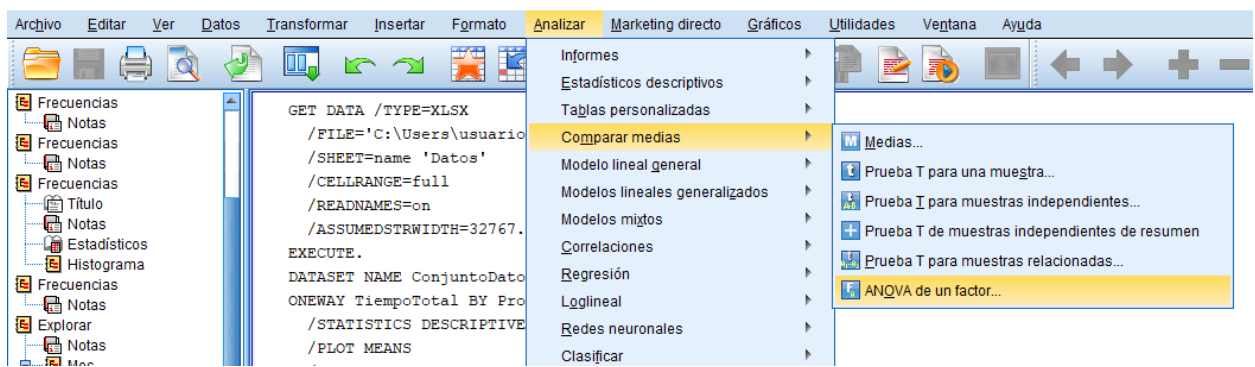


Figura 23. Análisis exploratorio. Adaptado de SPSS Statistics 23

Seleccionado ANOVA de un factor, se deben seleccionar las variables a analizar y su respectivo factor (en el caso que quiera hacer la comparación entre dos grupos) (ver figura 24).

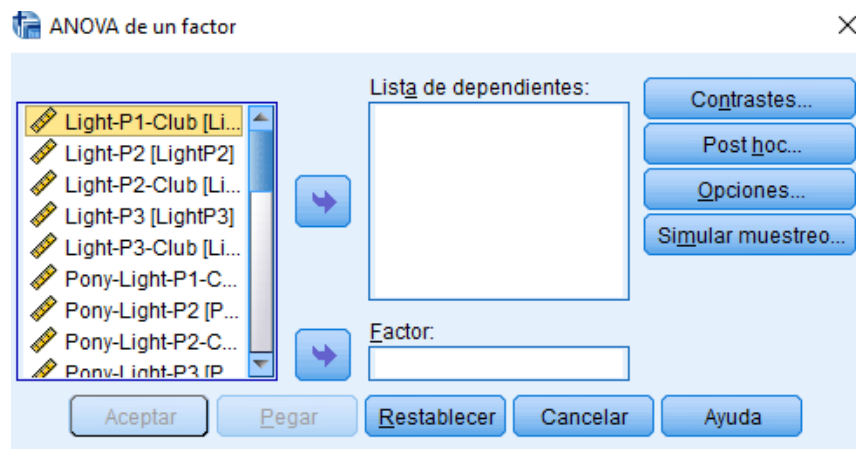


Figura 24. ANOVA de un factor. Adaptado de SPSS Statistics 23

En el botón “Opciones” se debe seleccionar las respectivas opciones de interés, para este caso, se seleccionan las opciones de Descriptivos, Prueba de homogeneidad de las varianzas y los respectivos Gráficos de medias (ver figura 25)

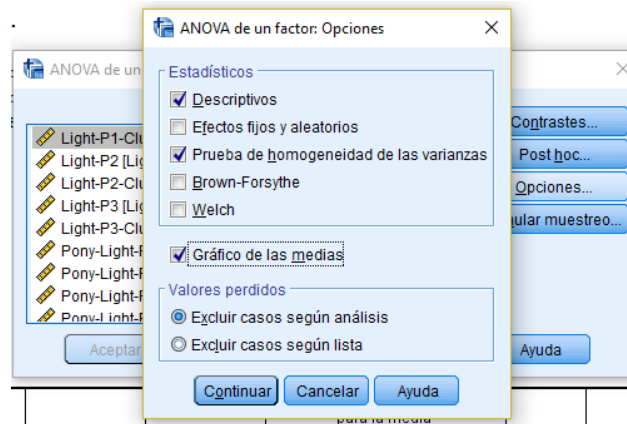


Figura 25. Opciones del ANOVA de un factor. Adaptado de SPSS Statistics 23

Una vez seleccionado los datos de interés y aceptando el análisis, se obtienen los resultados. Cabe recordar que se puede jugar con los datos de acuerdo con el conocimiento que se desea obtener.

```

/STATISTICS DESCRIPTIVES HOMOGENEITY
/PLOT MEANS
/MISSING ANALYSIS.
    
```

Unidireccional

[ConjuntoDatos1]

Descriptivos

TiempoTotal

	N	Media	Desviación estándar	Error estándar	95% del intervalo de confianza para la media		Mínimo	Máximo
					Límite inferior	Límite superior		
Agulla	122	56,50	74,691	6,762	43,11	69,89	1	685
Club	59	151,56	111,389	14,502	122,53	180,59	32	733
Light	155	124,57	106,175	8,528	107,72	141,42	8	629
Pilsen	131	55,17	47,805	4,177	46,90	63,43	3	239
Poker	116	58,40	49,092	4,558	49,37	67,43	3	307
Pony Malta	48	35,85	40,624	5,864	24,06	47,65	1	157
Total	631	80,61	86,264	3,434	73,87	87,35	1	733

Prueba de homogeneidad de varianzas

TiempoTotal

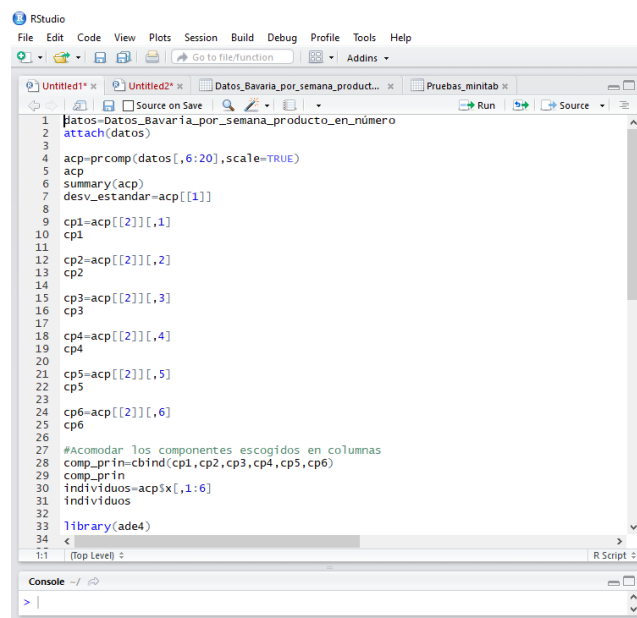
Estadístico de Levene	gl1	gl2	Sig.
11,834	5	625	,000

Figura 26. Resultados del ANOVA de un factor. Adaptado de SPSS Statistics 23

6.4.2 Paso a paso del análisis de componentes principales. Al aplicar esta técnica estadística se busca sintetizar la información, es decir, al presentar un gran número de variables, el objetivo será agruparlas para formar los nuevos componentes principales donde cada componente representa de forma lineal las variables que lo conforman. Además, estos nuevos componentes principales son independientes entre sí.

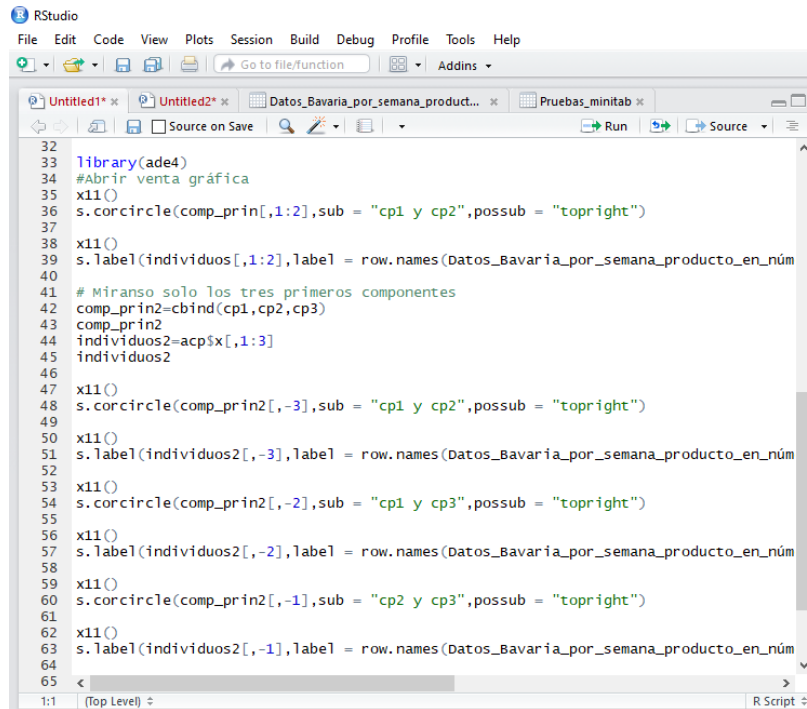
Para su desarrollo se utiliza el programa RStudio. Aquí se pueden observar todos los componentes principales resultantes y su porcentaje de participación en la varianza. Adicionalmente, se puede graficar y ver la relación de los individuos entre cada componente principal de modo que se puedan identificar individuos atípicos

Una vez se obtienen los resultados, cada componente contiene un porcentaje de participación de cada variable. El siguiente paso es observar las variables e identificar en cual componente presenta mayor participación en valor absoluto y así ir agrupando que variables corresponden a cada componente principal.



```
1 datos=Datos_Bavaria_por_semana_producto_en_numero
2 attach(datos)
3
4 acp=prcomp(datos[,6:20],scale=TRUE)
5 acp
6 summary(acp)
7 desv_estandar=acp[[1]]
8
9 cp1=acp[[2]][,1]
10 cp1
11
12 cp2=acp[[2]][,2]
13 cp2
14
15 cp3=acp[[2]][,3]
16 cp3
17
18 cp4=acp[[2]][,4]
19 cp4
20
21 cp5=acp[[2]][,5]
22 cp5
23
24 cp6=acp[[2]][,6]
25 cp6
26
27 #Acomodar los componentes escogidos en columnas
28 comp_prin=cbind(cp1,cp2,cp3,cp4,cp5,cp6)
29 comp_prin
30 individuos=acp$x[,1:6]
31 individuos
32
33 library(ade4)
```

Figura 27. Análisis en PCA. Adaptado de RStudio



```

32 library(ade4)
33 #Abrir venta gráfica
34 x11()
35 s.corcircle(comp_prin[,1:2],sub = "cp1 y cp2",possub = "topright")
36
37
38 x11()
39 s.label(individuos[,1:2],label = row.names(Datos_Bavaria_por_semana_producto_en_núm
40
41 # Miranso solo los tres primeros componentes
42 comp_prin2=cbind(cp1,cp2,cp3)
43 comp_prin2
44 individuos2=acp$x[,1:3]
45 individuos2
46
47 x11()
48 s.corcircle(comp_prin2[,-3],sub = "cp1 y cp2",possub = "topright")
49
50 x11()
51 s.label(individuos2[,-3],label = row.names(Datos_Bavaria_por_semana_producto_en_núm
52
53 x11()
54 s.corcircle(comp_prin2[,-2],sub = "cp1 y cp3",possub = "topright")
55
56 x11()
57 s.label(individuos2[,-2],label = row.names(Datos_Bavaria_por_semana_producto_en_núm
58
59 x11()
60 s.corcircle(comp_prin2[,-1],sub = "cp2 y cp3",possub = "topright")
61
62 x11()
63 s.label(individuos2[,-1],label = row.names(Datos_Bavaria_por_semana_producto_en_núm
64
65 <

```

Figura 28. Código para gráficas en PCA. Adaptado de RStudio

6.4.3 Paso a paso regresión lineal múltiple. El objetivo al aplicar esta técnica en este proyecto es ver la posible relación entre varias variables independientes (predictoras) y una variable dependiente. En este caso, se trabajaron con algunas variables categóricas transformadas en variables dicotómicas.

El programa utilizado es IBM SPSS, donde se siguen los siguientes pasos:

Primero se sigue la ruta Analizar > Regresión > Lineales (ver figura 29).

Una vez se muestre el cuadro, se debe seleccionar las variables de interés para el estudio. En la opción de método, vale la pena ejecutar primero con la opción “Intro” para ver el comportamiento de las variables predictoras con la variable dependiente. Luego, cambiar la opción “Intro” por “Paso a paso” donde el programa crea modelos con las variables significativas y las otras son excluidas.

En la opción de “Estadísticos”, se debe seleccionar los ítems de descriptivos, el diagnóstico de colinealidad, las respectivas correlaciones y el estadístico R cuadrado para tomar decisiones justificadas.

En “Gráficos” se tiene la opción de seleccionar algunas variables predictoras respecto a la variable dependiente inicial (ver figura 30).

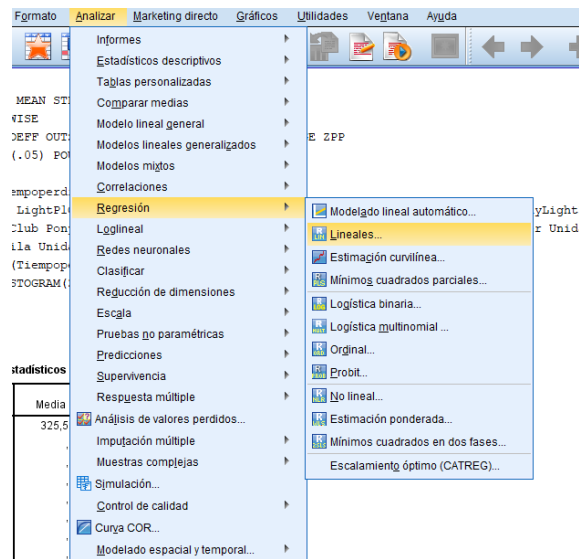


Figura 29. Análisis de análisis de regresión múltiple. Adaptado de SPSS Statistics 23

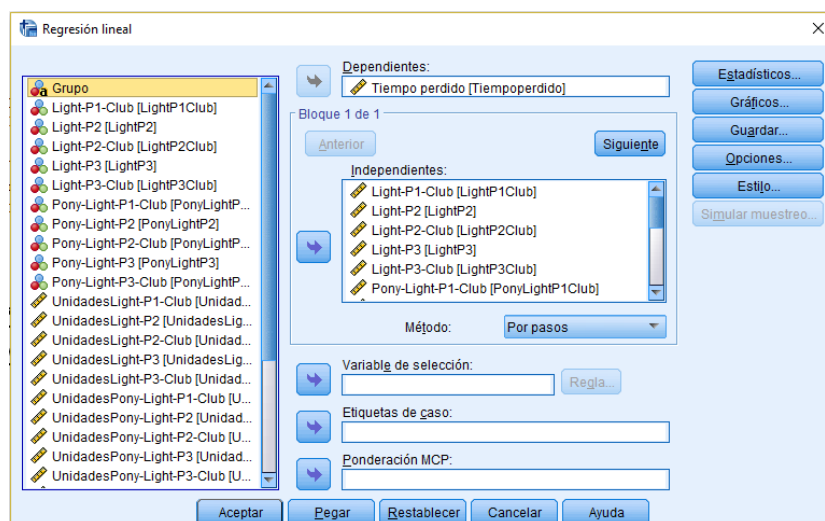


Figura 30. Selección de variables en regresión múltiple. Adaptado de SPSS Statistics 23

De acuerdo con los resultados y definido el modelo de predicción apropiado, se procede a calcular la variable independiente predicha para después realizar una prueba T para las dos muestras relacionadas, es decir, entre el valor real y el valor predicho. Con base a esa comparación, se concluye si el modelo predictivo representa la variable independiente (ver figura 32).

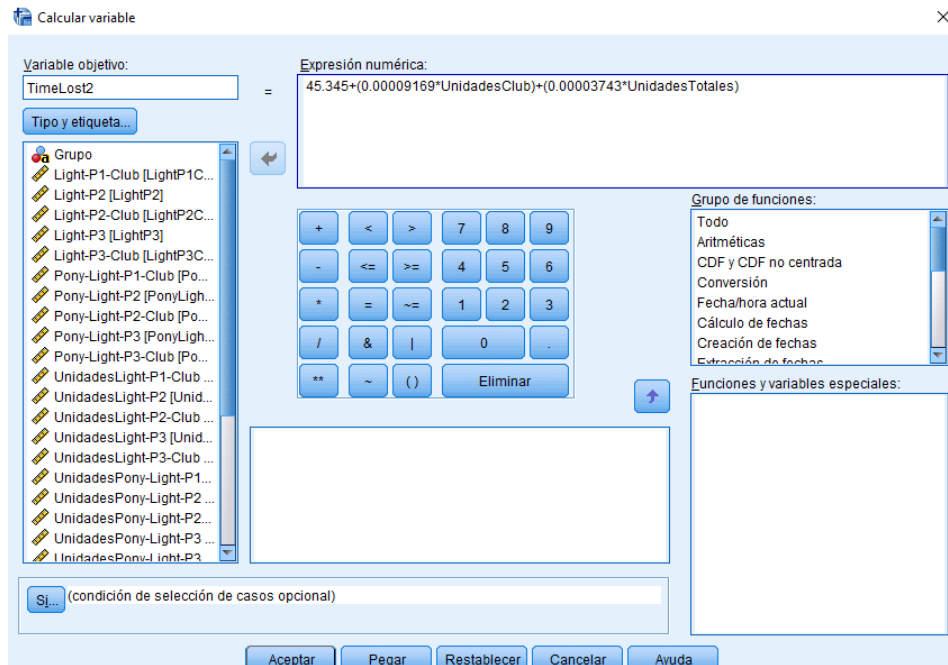


Figura 31. Cálculo de variable predictorica. Adaptado de SPSS Statistics 23



Figura 32. Análisis de prueba T para muestras relacionadas. Adaptado de SPSS Statistics 23

Para desarrollar el análisis de regresión múltiple se contaron con diferentes dificultades cuya solución se encuentra a continuación:

- El número de individuos no era suficiente para la cantidad de grupos que salieron. Para solucionar esto, se crearon los grupos con la letra P acompañado de su respectivo número (lo relacionado a la P y el número es explicado en la sección 6.3.3), esto fue posible dado que estos tres productos (pilsen, poker y águila) presenta una media similar, es decir, el orden de su elaboración no afecta el tiempo final perdido. Hecho esto, se satisface el concepto de que mínimo por cada grupo deben existir diez (10) individuos. Finalmente resultaron 10 grupos con un total de 140 individuos
- El siguiente problema era como utilizar estos grupos para el modelo de regresión. Para solucionar esto, se decide crear igual número de variables con los grupos para ser transformadas en variables dummy o categóricas (ver figura 33). Con esto, se pretende identificar cuales grupos son los más significativos para la investigación y una vez hallados, usarlos para calcular el respectivo modelo de predicción.

Grupos de productos	Light-P1-Club	Light-P2	Light-P2-Club	Light-P3	Light-P3-Club	Pony-Light-P1-Club	Pony-Light-P2	Pony-Light-P2-Club	Pony-Light-P3	Pony-Light-P3-Club
Light-P2	0	1	0	0	0	0	0	0	0	0
Light-P3	0	0	0	1	0	0	0	0	0	0
Light-P3	0	0	0	1	0	0	0	0	0	0
Light-P3	0	0	0	1	0	0	0	0	0	0
Pony-Light-P1-Club	0	0	0	0	0	1	0	0	0	0
Light-P3	0	0	0	1	0	0	0	0	0	0
Light-P2-Club	0	0	1	0	0	0	0	0	0	0
Pony-Light-P3	0	0	0	0	0	0	0	0	1	0
Light-P3	0	0	0	1	0	0	0	0	0	0
Pony-Light-P3	0	0	0	0	0	0	0	0	0	1
Light-P2-Club	0	0	1	0	0	0	0	0	0	0
Light-P2-Club	0	0	1	0	0	0	0	0	0	0
Pony-Light-P3	0	0	0	0	0	0	0	0	1	0
Light-P2-Club	0	0	1	0	0	0	0	0	0	0
Light-P3	0	0	0	1	0	0	0	0	0	0
Pony-Light-P2	0	0	0	0	0	0	1	0	0	0
Light-P1-Club	1	0	0	0	0	0	0	0	0	0
Light-P2	0	1	0	0	0	0	0	0	0	0
Pony-Light-P2-Club	0	0	0	0	0	0	0	1	0	0
Light-P2	0	1	0	0	0	0	0	0	0	0
Light-P2	0	1	0	0	0	0	0	0	0	0
Pony-Light-P2-Club	0	0	0	0	0	0	0	1	0	0
Light-P3	0	0	0	1	0	0	0	0	0	0
Light-P3	0	0	0	1	0	0	0	0	0	0
Pony-Light-P2-Club	0	0	0	0	0	0	0	1	0	0

Figura 33. Creación de variables dummy. Adaptado de Apéndice D

- Otra dificultad era saber si utilizar las unidades elaboradas como una sola variable o separar esas unidades por cada producto y crear igual número de variables. Se analizaron los siguientes casos: (1) una sola variable como unidades totales, (2) seis variables (unidades pony, unidades light, unidades pilsen, unidades poker, unidades águila y unidades club) cada una con su respectiva cantidad de productos y (3) dejando las seis variables de las unidades de cada producto y la variable que engloba la cantidad total. Hecho el análisis bajo el programa SPSS, se concluye aplicar el caso 3 ya que una variable (unidades club) es significativa para la investigación al igual que las unidades totales.
- Con el uso de las variables dummy y aplicando el modelo “Intro” que brinda el programa SPSS se pretende identificar los grupos significativos, es decir, cuando se diseña la producción en el orden que indican los grupos seleccionados, el tiempo perdido será mayor. Para la creación del modelo se modifica el modelo a “paso a paso” con el objetivo que trabaje solamente con las variables significativas y me excluya las otras.
- El dilema del modelo creado es observar si pronostica de forma correcta el tiempo perdido para ciclos de producción del año 2018. Para esto, se acudió a la empresa para la respectiva toma de los datos del año 2018, tomar algunas semanas al azar y comparar el tiempo real perdido contra el pronosticado por el modelo. Este resultado se puede observar en el siguiente capítulo (parte 7.3).
- Con base a la revisión de literatura, no se encuentran trabajos relacionados a la aplicación de técnicas de minería de datos en la industria cervecera, especialmente el área de embotellado. Debido a esto, se encontraron dificultades al momento de saber cuál era la técnica apropiada a utilizar.

7. Resultados

Para presentar los resultados, se seguirá el mismo orden del capítulo anterior, es decir, inicia con un análisis previo de los datos, el tiempo perdido por cada producto, cuales meses son las más significativos, las máquinas que representan mayor pérdida de tiempo, etc. Continúa con el PCA para finalizar con el modelo de predicción del tiempo perdido en cada ciclo de producción.

7.1 Resultados análisis previo de los datos

En la figura 34 se observa la media de pérdida de tiempo de cada producto cuando es elaborado. Se identifica el producto club con una media de 151 minutos, como el que presenta mayor pérdida. En segundo puesto se encuentra el producto light con una media de 124 minutos. Los productos águila, pilsen y poker presenta una media de 56, 55 y 58 minutos respectivamente. De acuerdo con la figura 35, con un p valor de 0.984, se considera similar la media de estos tres productos. Por tal motivo en la actualidad estos tres productos son elaborados en forma sucesiva.

Cabe nombrar que el producto Pony es el que presenta mayor media de tiempo perdido en los tres (3) años e igualmente es el producto del cual se elaboran menos unidades. Con el producto club sucede lo contrario, es el penúltimo producto en unidades elaboradas, pero es el primer en tiempo perdido, de ahí que se define como el producto crítico del proceso. Por tal motivo, cuando en el ciclo de elabora club, este va siempre de último.

Tabla 5.

Lista de máquinas críticas

MÁQUINA	TIEMPO PERDIDO EN LOS 3 AÑOS [MIN]
Lavadora de botellas	9546
Transportador de botellas	8966
Envasadora	8587
Etiquetadora	4873

Las siguientes gráficas representan el tiempo perdido de estas máquinas en los tres años respecto a cada producto.

Observando desde la figura 36 a la gráfica 39 se identifica que las máquinas presentan la misma tendencia de pérdida por producto al igual que la figura 34. El producto club lidera la pérdida en las cuatro máquinas, seguido del producto light, como era de esperarse. Aquí se ratifica que los productos pilsen, poker y águila presenta medias similares ya que los tres presentan pérdidas constantes en las cuatro máquinas.

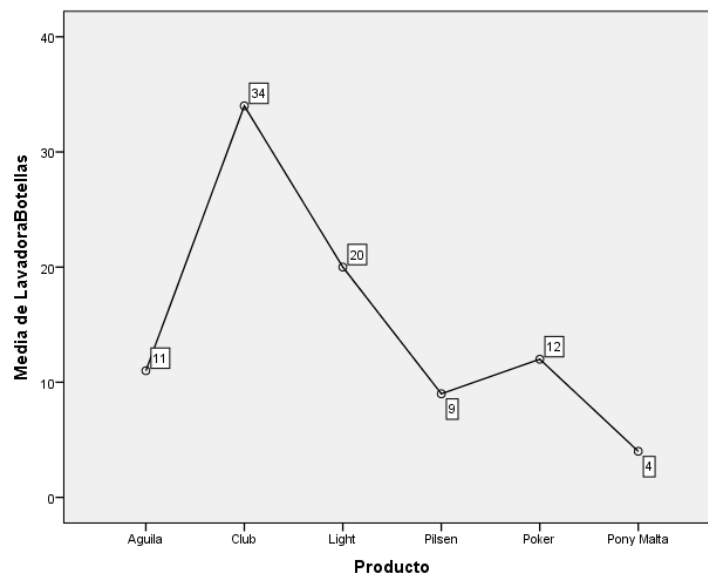


Figura 36. Gráfica producto vs media de tiempo perdido en lavadora de botellas. Adaptado de SPSS Statistics 23

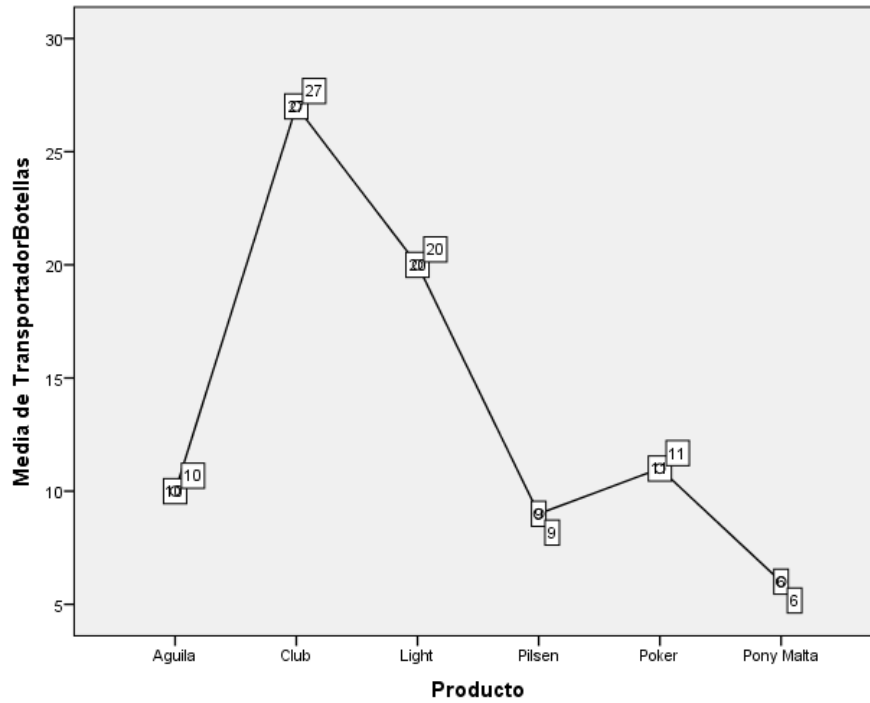


Figura 37. Gráfica producto vs media de tiempo perdido en transportado de botellas. Adaptado de SPSS Statistics 23

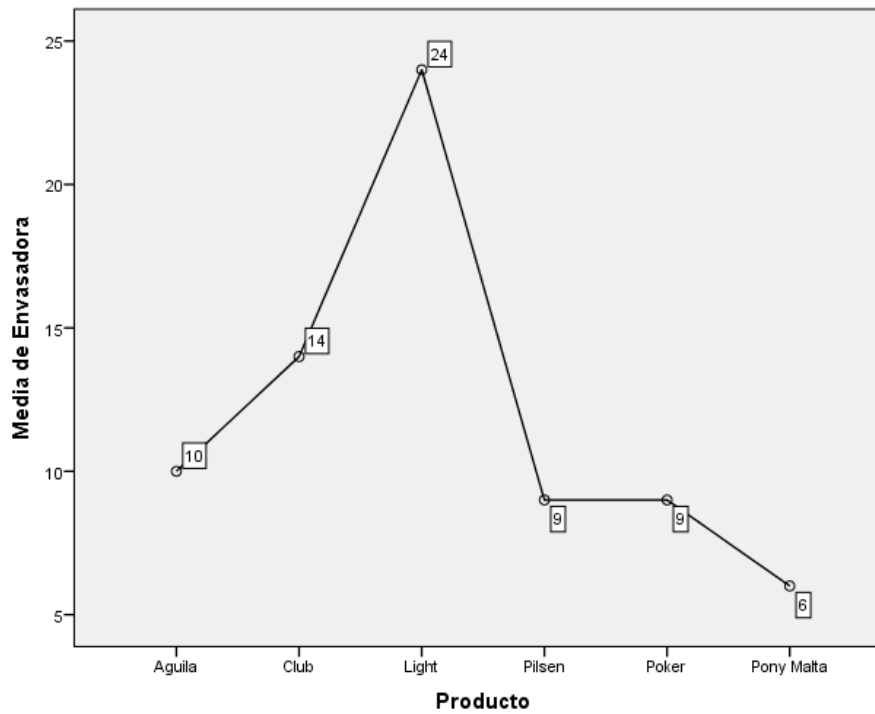


Figura 38. Gráfica producto vs media de tiempo perdido en envasadora. Adaptado de SPSS Statistics 23

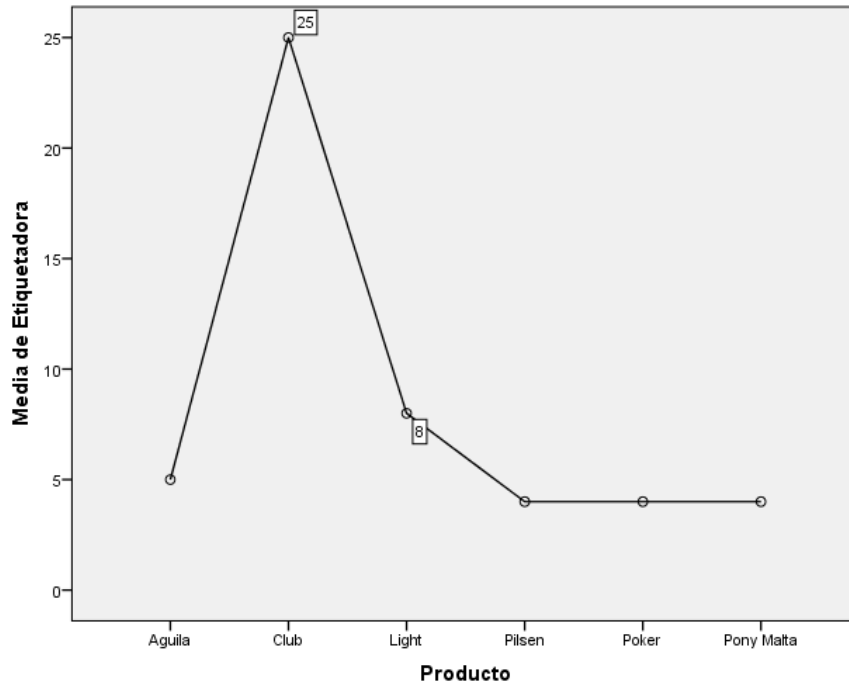


Figura 39. Gráfica producto vs media de tiempo perdido en etiquetadora. Adaptado de SPSS Statistics 23

Ya analizado por productos, ahora observar el comportamiento por meses y ver cuales con los que representan mayor pérdida de tiempo. El ANOVA de la figura 40, con un nivel de significancia de 0.661 se acepta la hipótesis nula, es decir, la media del tiempo perdido por mes es similar.

$H_0 > 0.05$: la media del tiempo perdido por meses es igual

$H_1 < 0.05$: la media del tiempo perdido por meses es diferente

ANOVA

TiempoTotal

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Entre grupos	64071,339	11	5824,667	,780	,661
Dentro de grupos	4624076,756	619	7470,237		
Total	4688148,095	630			

Figura 40. ANOVA de tiempo perdido por meses. Adaptado de SPSS Statistics

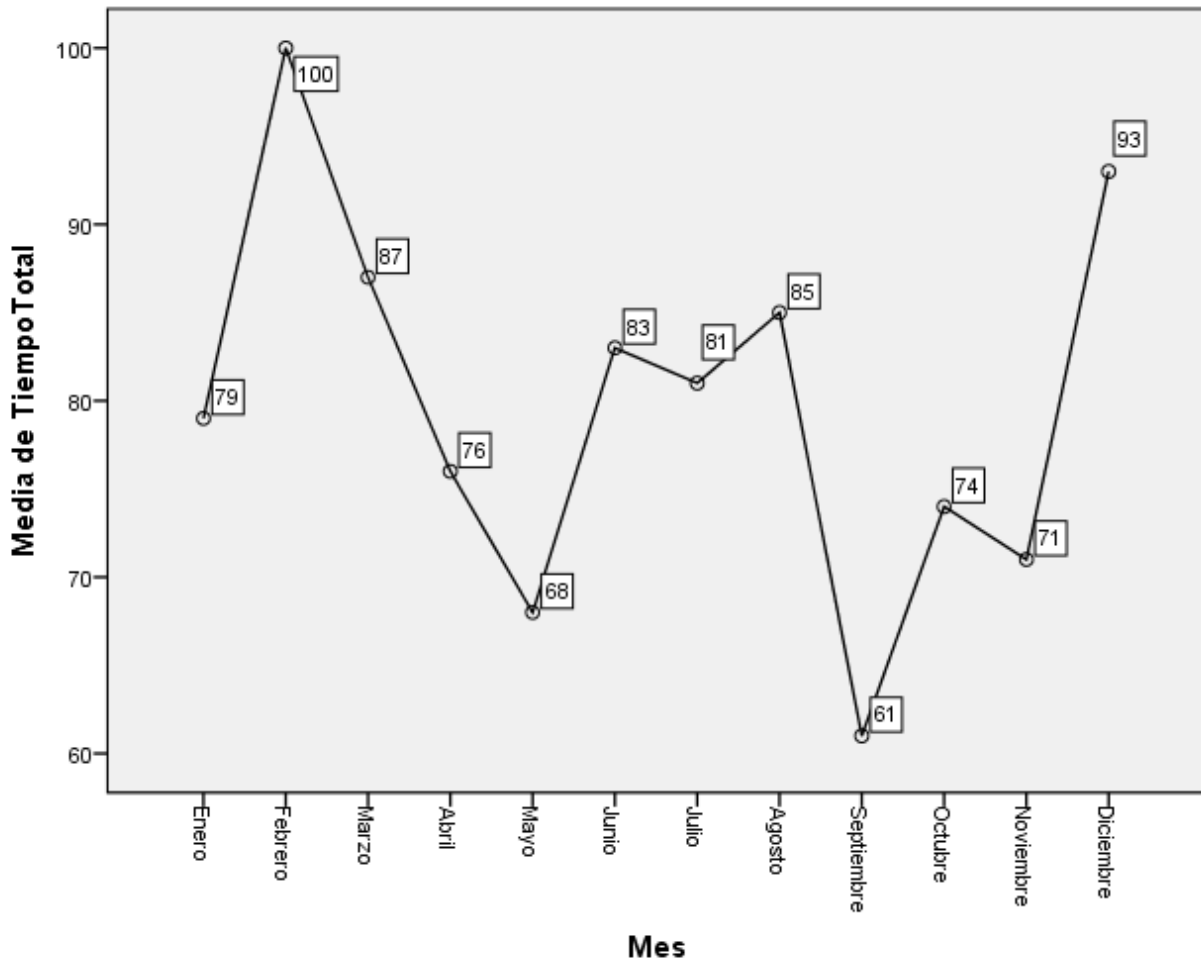


Figura 41. Gráfica mes vs media de tiempo total perdido. Adaptado de SPSS Statistics

Con base a la figura 41, los meses de febrero, diciembre y marzo con 100, 93 y 87 minutos respectivamente, son lo más significativos en relación con el tiempo perdido en los tres años. Se destaca una tendencia constante entre los meses de junio, julio y agosto con valores de 83, 81 y 85 minutos respectivamente. Septiembre es el mes que menor pérdida de tiempo presenta.

Ahora, el siguiente paso es observar el comportamiento por mes de las máquinas críticas y ver si presentan la tendencia de la figura 41 o identificar que máquinas son representativas de un mes en particular.

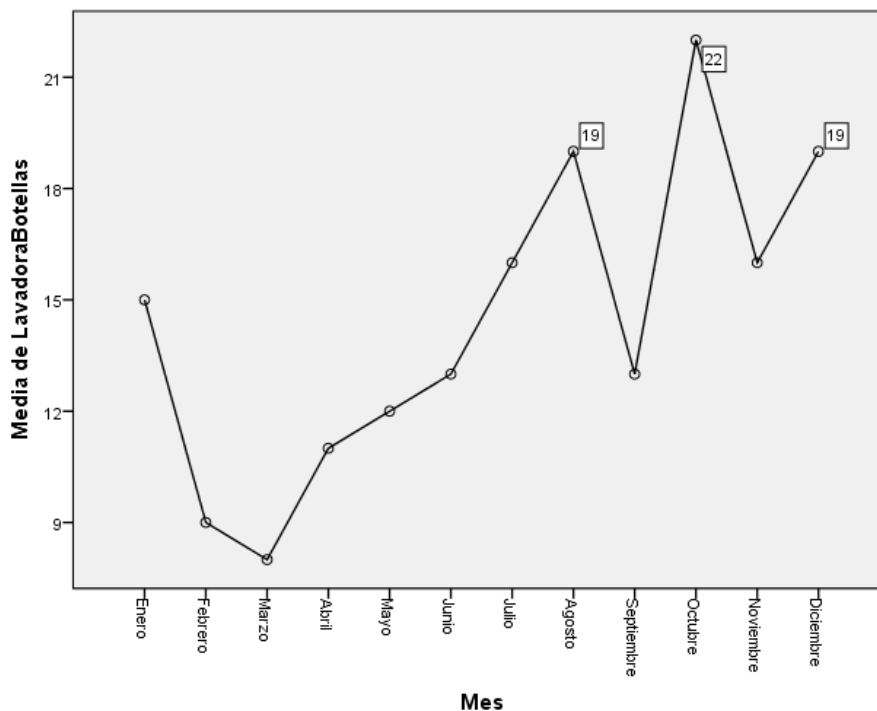


Figura 42. Gráfica mes vs media de tiempo perdido en lavadora de botellas. Adaptado de SPSS Statistics 23

Analizando la figura 42 a la figura 45, se evidencia que las máquinas presentan punto elevados en los meses de febrero, marzo y diciembre. Se puede concluir el aporte de cada máquina al mes mediante la tabla

Tabla 6.
Meses críticos de cada máquina

MÁQUINA	MESES
Lavadora de botellas	Agosto, octubre y diciembre
Transportador de botellas	Febrero y agosto
Envasadora	Febrero, diciembre y marzo
Etiquetadora	Marzo y febrero

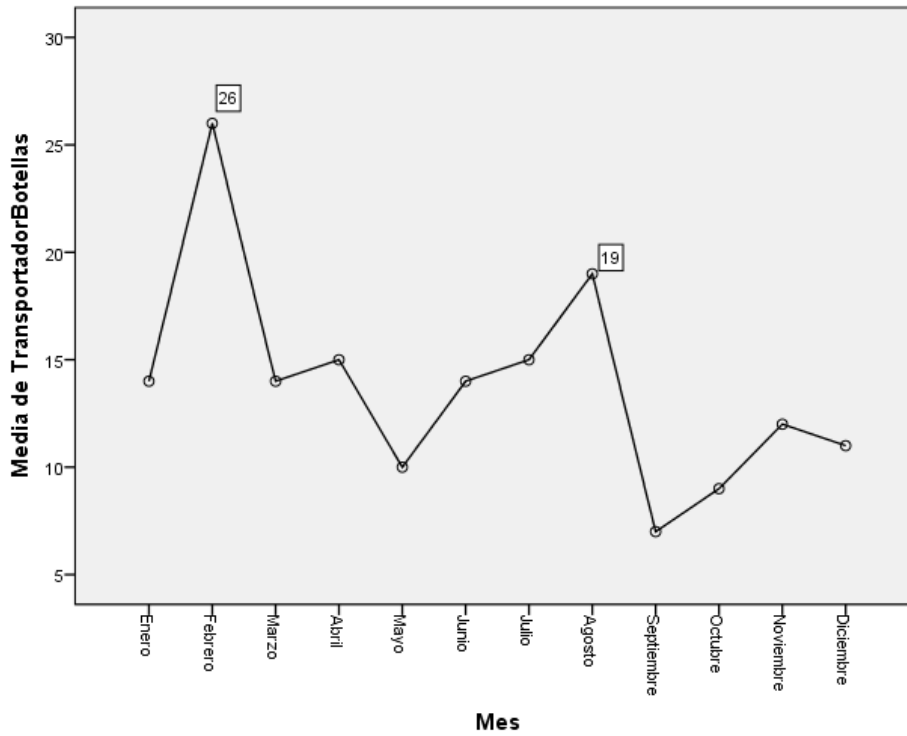


Figura 43. Gráfica mes vs media de tiempo perdido en transportador de botellas. Adaptado de SPSS Statistics 23

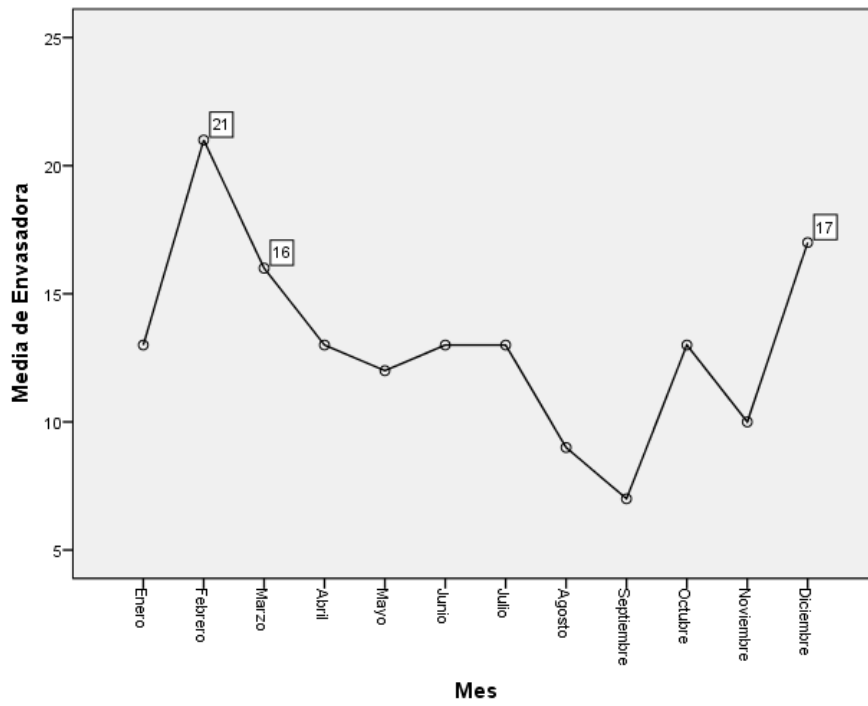


Figura 44. Gráfica mes vs media de tiempo perdido en envasadora. Adaptado de SPSS Statistics 23

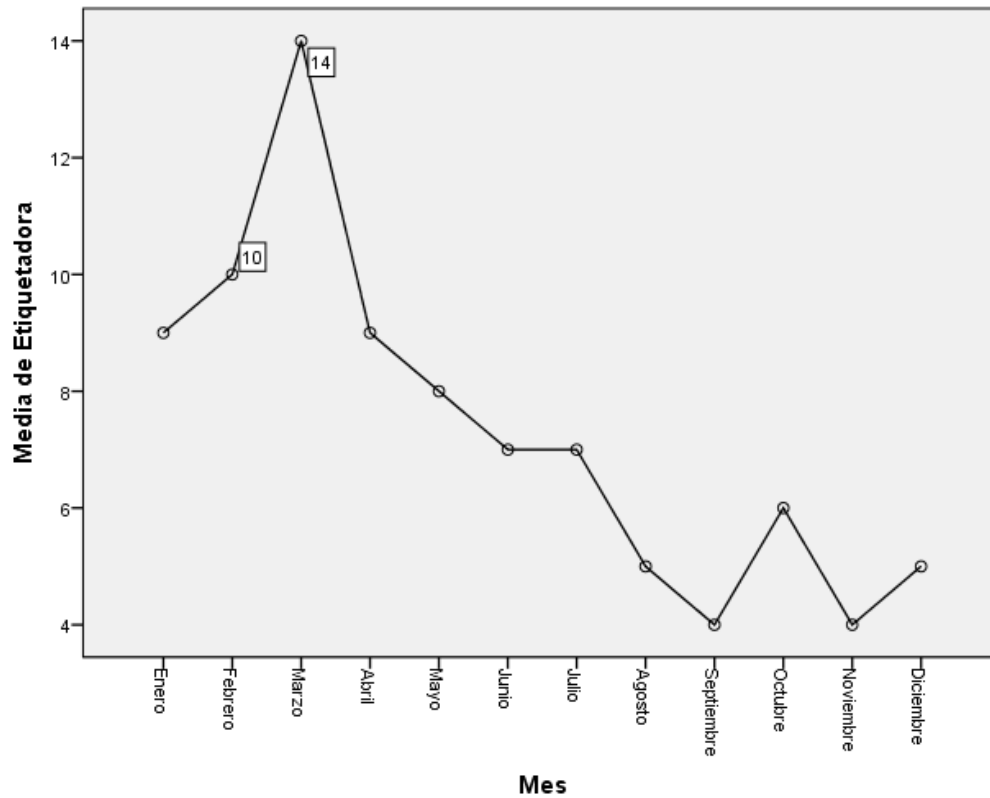


Figura 45. Gráfica mes vs media de tiempo perdido en etiquetadora. Adaptado de SPSS Statistics 23

Analizado por producto y por mes, ahora se analiza por semana. Es importante para ver si los resultados siguen concordando porque respecto a producto, máquina y mes vienen siendo correctos.

ANOVA

TiempoTotal

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Entre grupos	347684,309	52	6686,237	,890	,691
Dentro de grupos	4340463,787	578	7509,453		
Total	4688148,095	630			

Figura 46. ANOVA de tiempo perdido por semana. Adaptado de SPSS Statistics 23

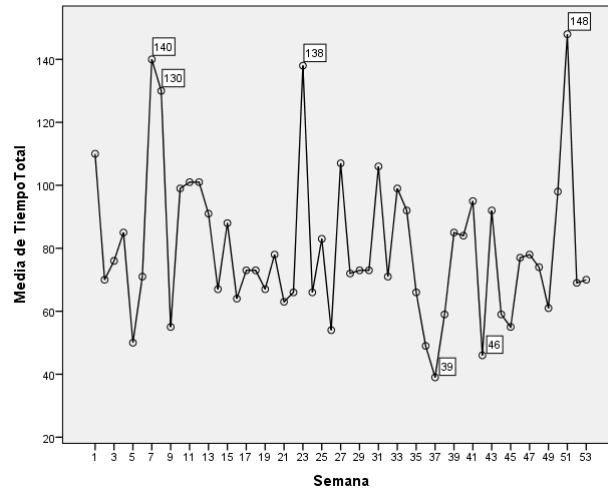


Figura 47. Gráfica semanas vs media de tiempo total perdido. Adaptado de SPSS Statistics 23

Se presentan picos elevados en las semanas que corresponden a los meses donde también se presentan valores altos de pérdida de tiempo. Las semanas 5 y 7 que corresponden al mes de febrero, la semana 23 que corresponde a mayo y la semana 52 que es del mes de diciembre. Estos resultados eran de esperarse debido al análisis realizado por mes.

Ahora miremos cómo se comportan las máquinas críticas en tiempo perdido por semana.

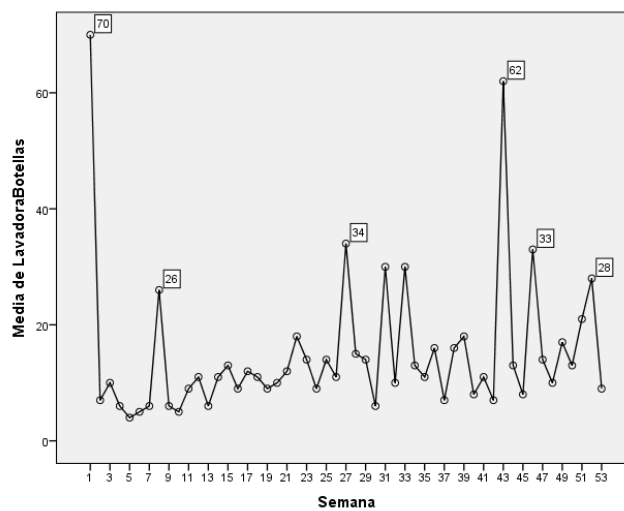


Figura 48. Gráfica semana vs media de tiempo perdido en lavadora de botellas. Adaptado de SPSS Statistics 23

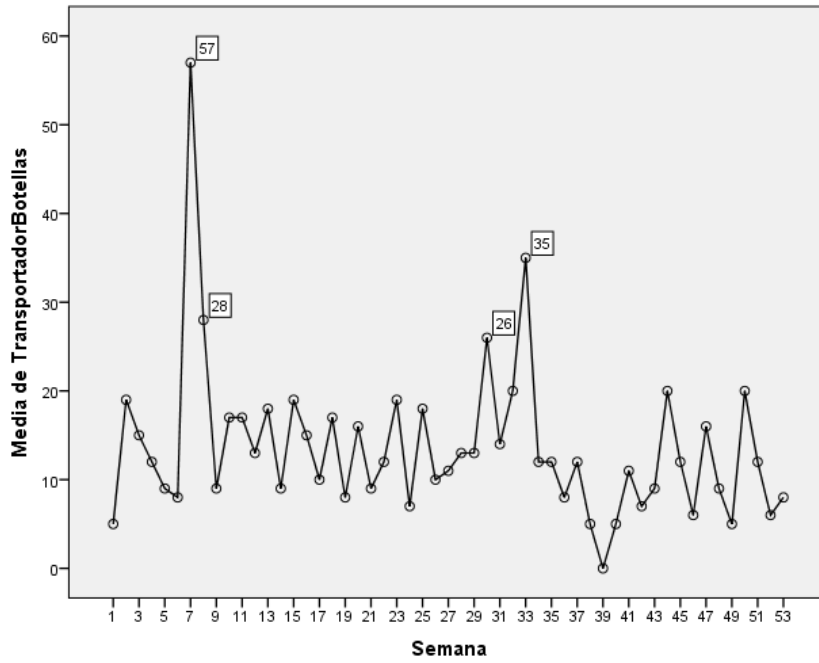


Figura 49. Gráfica semana vs media de tiempo perdido en transportador de botellas. Adaptado de SPSS Statistics 23

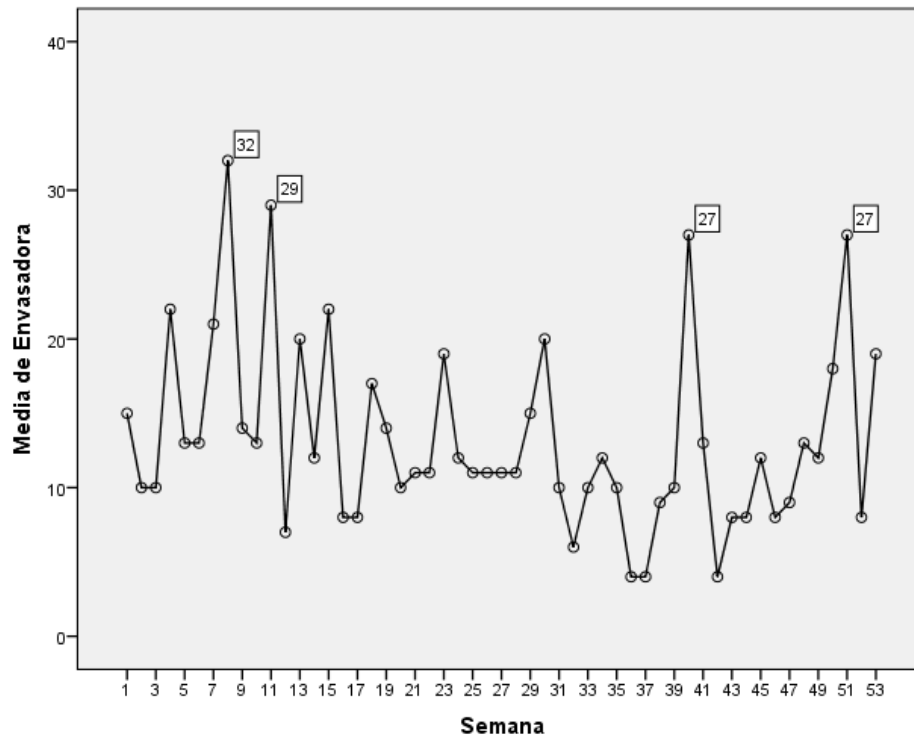


Figura 50. Gráfica semana vs media de tiempo perdido en envasadora. Adaptado de SPSS Statistics 23

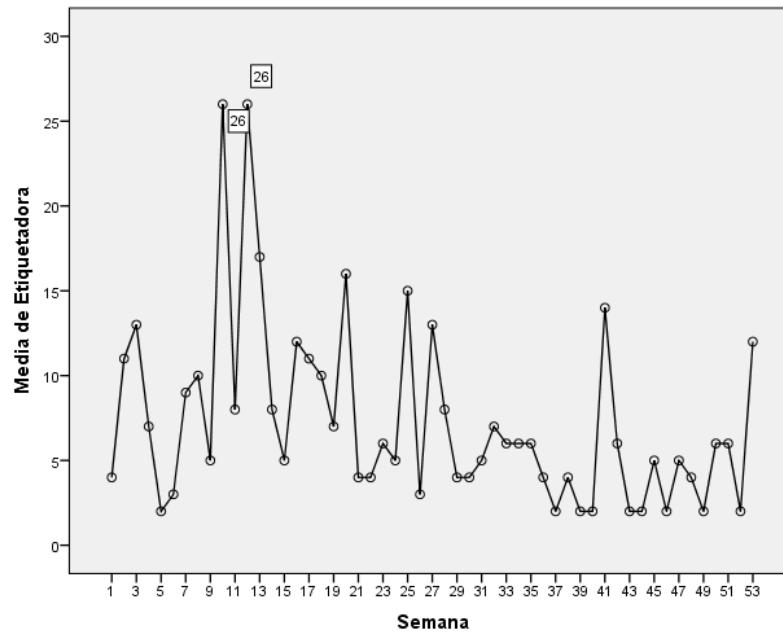


Figura 51. Gráfica semana vs media de tiempo perdido en etiquetadora. Adaptado de SPSS Statistics 23

De acuerdo con las gráficas se observa puntos elevados en las semanas del mes de febrero de las máquinas transportador de botellas, envasadora y etiquetadora. Para el mes de diciembre se presenta pérdidas elevadas en las máquinas lavadora de botellas y envasadora. Adicional, se identifica que las semanas que corresponden al mes de septiembre (mes con menos pérdida en tiempo) presentan valores bajos, lo que era de esperarse.

7.2 Resultados análisis de componentes principales

Determinante = 0.326

Prueba de KMO y Bartlett		
Medida Kaiser-Meyer-Olkin de adecuación de muestreo		,573
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	699,072
	gl	105
	Sig.	,000

Figura 52. Prueba KMO y Bartlett. Adaptado de SPSS Statistics 23

Los determinantes próximos a cero están indicando que las variables utilizadas están linealmente relacionadas, lo que significa que aplicar PCA es una técnica correcta para aplicar a esas variables. En este caso, el determinante tiene un valor de 0.326, es decir, se puede aplicar la técnica. Presenta un KMO de 0.573 el cual es bajo, pero este acompañado de un pvalor menor que 0.05, es decir, se acepta la hipótesis nula la cual indica que se pueda aplicar el análisis de componentes principales.

Hecho el análisis, se obtiene el gráfico de sedimentación (ver figura 53). Este gráfico muestra la varianza asociada a cada factor y de acuerdo con esto, se determina cuantos factores deben retenerse. Típicamente se muestran dos rupturas entre la pronunciada pendiente de los factores más importante.

Existe otra opción para seleccionar los factores que es utilizando el criterio de kaiser, que consiste en retener aquellos factores cuyo autovalor sea mayor a 1.

Aplicado el análisis de componentes principales se obtienen los siguientes resultados:

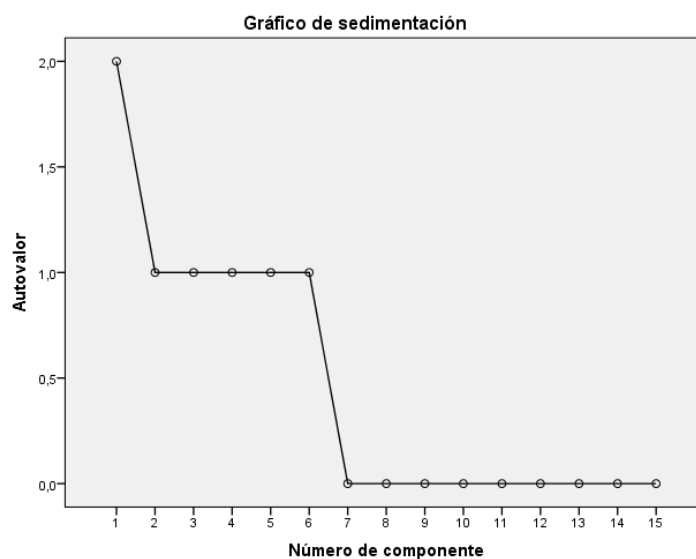


Figura 53. Gráfico de sedimentación. Adaptado de SPSS Statistics 23

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Standard deviation	1.4326	1.2509	1.12731	1.06461	1.04419	1.0025	0.99215	0.9759
Proportion of Variance	0.1368	0.1043	0.08472	0.07556	0.07269	0.0670	0.06562	0.0635
Cumulative Proportion	0.1368	0.2412	0.32587	0.40143	0.47412	0.5411	0.60674	0.6702
	PC9	PC10	PC11	PC12	PC13	PC14	PC15	
Standard deviation	0.96950	0.93372	0.90623	0.87108	0.76984	0.74020	0.6435	
Proportion of Variance	0.06266	0.05812	0.05475	0.05059	0.03951	0.03653	0.0276	
Cumulative Proportion	0.73290	0.79102	0.84577	0.89636	0.93587	0.97240	1.0000	

Figura 54. Componentes principales. Adaptado de RStudio.

De acuerdo con el criterio de kaiser, se seleccionan los seis (6) primeros componentes

	PC1	PC2	PC3	PC4	PC5	PC6
Desv. Est	1.4326	1.2509	1.12731	1.06461	1.04419	1.0025
Varianza en %	0.1368	0.1043	0.08472	0.07556	0.07269	0.0670
Acumulativa	0.1368	0.2412	0.32587	0.40143	0.47412	0.5411

	PC1	PC2	PC3	PC4
Depaletizadora	0.070611011	-	0.051088975	0.19316770
Desempacadora	0.219269106	0.478079441	-	0.30547758
LavadoraBotellas	0.135142258	0.006364293	0.26072057	0.2686
Insp. Botellas Vacías	0.141209259	0.002143089	0.32462107	-0.04899
Envasadora	0.271346808	0.003858167	0.37550663	-0.2249
Fechador	-0.009372757	0.057358767	0.06747235	0.36371
Pasteurizador	0.310483197	-0.265851267	-0.11903911	-0.18087
Etiquetadora	0.311791868	0.005006273	0.30993968	-0.32109
Insp. Prod. Terminado	0.162362281	-0.043369265	0.27406469	-0.14712
LavadoraCajas	0.202611888	0.404423820	0.02952353	0.49169
Empacadora	0.254030379	0.533657627	-0.26610183	-0.18861

InspectorCajaLlena	0.387308050	-0.348453461	-0.39696393	0.18724
Paletizadora	0.061590948	0.061620181	0.11923102	0.01271
TransportadorBotellas	0.452265920	-0.324827205	-0.25764334	0.0515
TransportadorCanastas	0.388729703	0.117551855	0.23506425	0.37747

	PC5	PC6
Depaletizadora	0.296455175	-0.6236353042
Desempacadora	0.077112926	-0.1256954519
LavadoraBotellas	-0.127050030	-0.1766376502
Insp. Botellas Vacías	0.192340098	-0.0635119758
Envasadora	0.312426572	0.0264621684
Fechador	0.224057758	-0.3584358196
Pasteurizador	-0.025762424	0.0340929691
Etiquetadora	-0.245658483	-0.0187530620
Insp. Prod. Terminado	-0.539383147	0.1712287668
LavadoraCajas	-0.137095163	-0.0293878601
Empacadora	-0.009977162	-0.0003832578
InspectorCajaLlena	0.063865074	0.0082436130
Paletizadora	0.572675369	0.6061685266
TransportadorBotellas	0.043927787	-0.0780207711
TransportadorCanastas	-0.020326928	0.1599249037

El resultado anterior permite identificar las máquinas que corresponden a cada factor principal. Esto se hace verificando en que componente la máquina presente mayor participación en valor

absoluto (valores marcados en rojo). Hecho esto, se obtienen los siguientes componentes principales (ver tabla 7)

Identificado las máquinas que corresponden a los componentes principales y relacionándolos con la distribución actual de cada máquina en la línea de producción en el área de embotellado de Bavaria sede Bucaramanga, se evidencia que las máquinas que están contiguas se agrupan formando las componentes (ver figura 55).

Tabla 7.
Componentes principales seleccionados

COMPONENTE	MÁQUINAS QUE LO CONFORMAN
CP1	Pasteurizador, etiquetadora, transportador de botellas y transportador de canastas
CP2	Empacadora y desempacadora
CP3	Envasadora, inspector de botellas vacías e inspector de caja llena
CP4	Lavadora de botellas y lavadora de cajas
CP5	Inspector de producto terminado
CP6	Paletizadora y depaletizadora

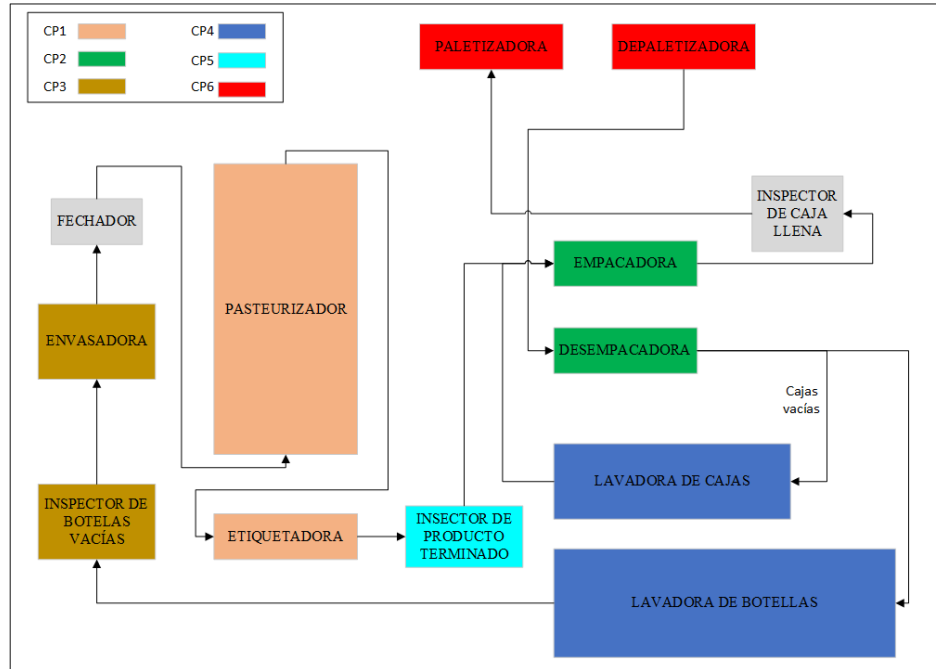


Figura 55. Distribución de las componentes en el layout actual de la línea de producción.

Ahora, se analiza la relación entre los primeros tres (3) componentes principales:

CP1 y CP2

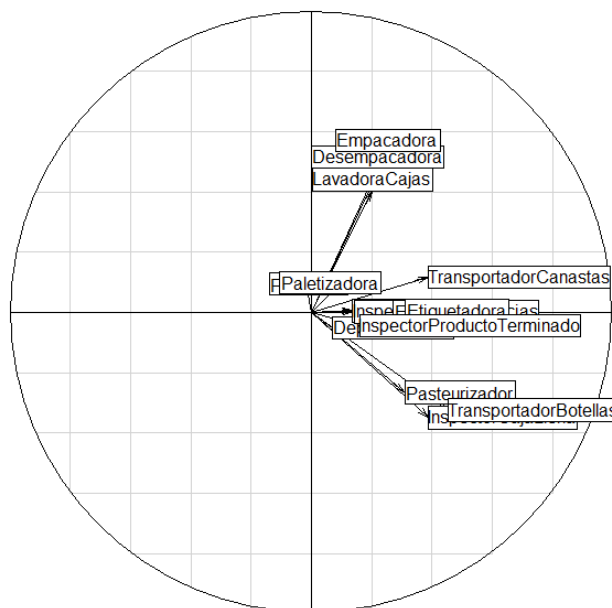


Figura 56. Ubicación de máquinas respecto a PC1 y PC2. Adaptado de RStudio.

En la figura 56 se observa la distribución de las máquinas respecto al componente PC1 y PC2. Se identifica claramente como el componente 2 representa la empacadora y desempacadora e igualmente para el componente 1 que representa el pasteurizador, etiquetadora y los transportadores.

En la figura 57 donde se gráfica los individuos para ver la relación respecto al componente PC1 y PC2, se identifican 4 individuos característicos, el 444, 87, 197 y 121. Es curioso observar que el 444 y 121 corresponden a producto club y el 197 y 121 corresponden al producto light; estos dos productos son lo que representan mayor pérdida de tiempo en sus ciclos de producción.

El individuo 87 presenta tiempo perdido en las máquinas del componente 2, es decir, 381 minutos en la empacadora y 112 minutos en la desempacadora.

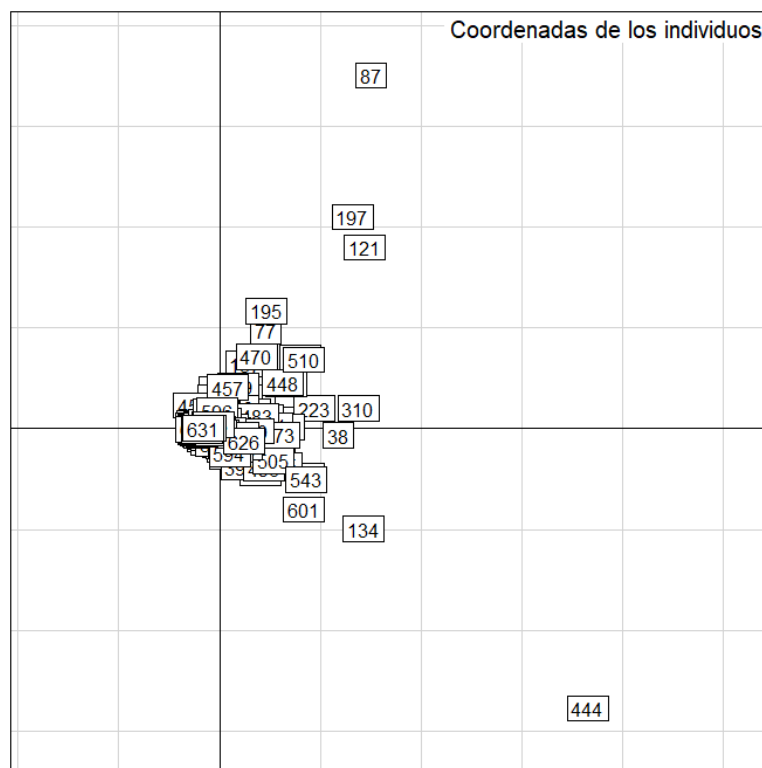


Figura 57. Coordenadas de los individuos respecto al PC1 y PC2. Adaptado de RStudio

El individuo 444 presenta tiempo perdido en las máquinas que corresponden al componente 1, en este caso, a las máquinas pasteurizador y transportador de botellas, 113 y 497 minutos respectivamente.

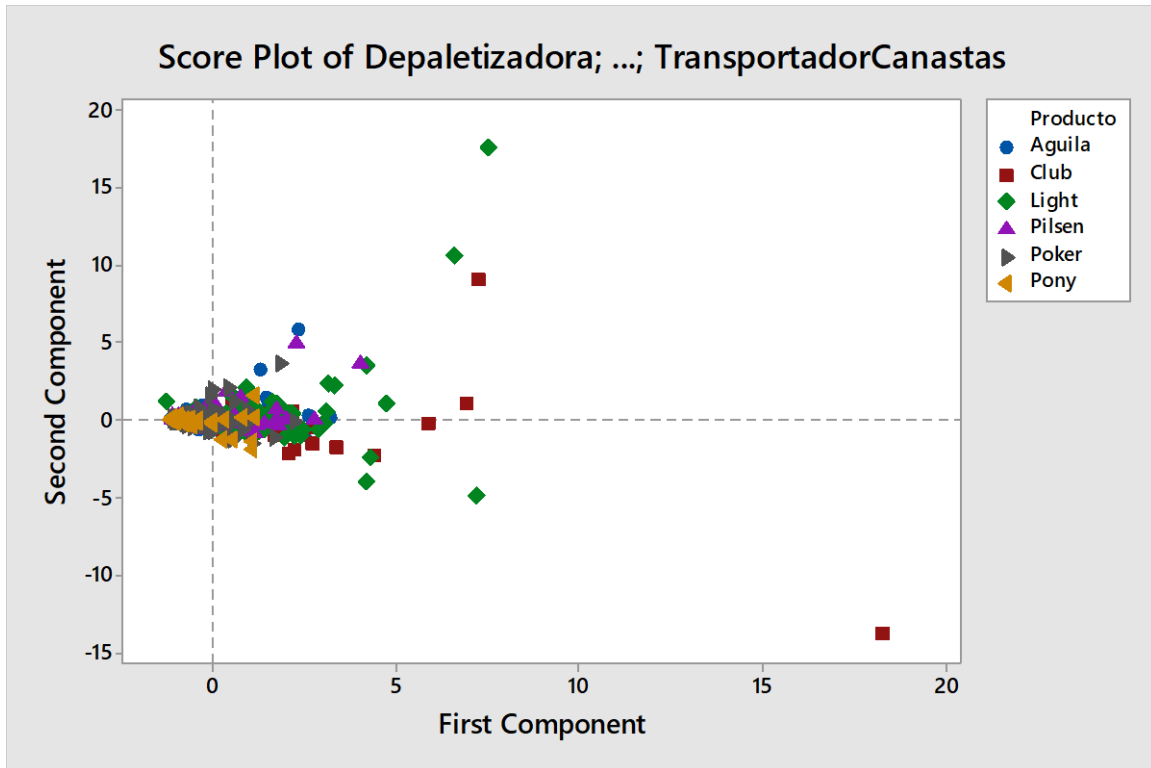


Figura 58. Coordenadas de los productos de cada individuo respecto a CP1 y CP2. Adaptado de Minitab

CP1 Y CP3

En la figura 59 se observa la distribución de las máquinas respecto al componente PC1 y PC2. Se identifica claramente como el componente 3 representa la envasadora y el inspector de botellas vacías, igualmente para el componente 1 que representa el pasteurizador, etiquetadora y los transportadores.

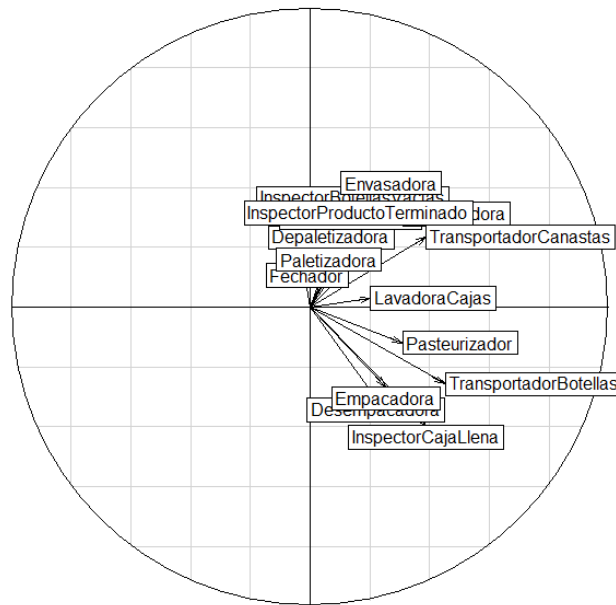


Figura 59. Ubicación de máquinas respecto a CP1 y CP3. Adaptado de RStudio

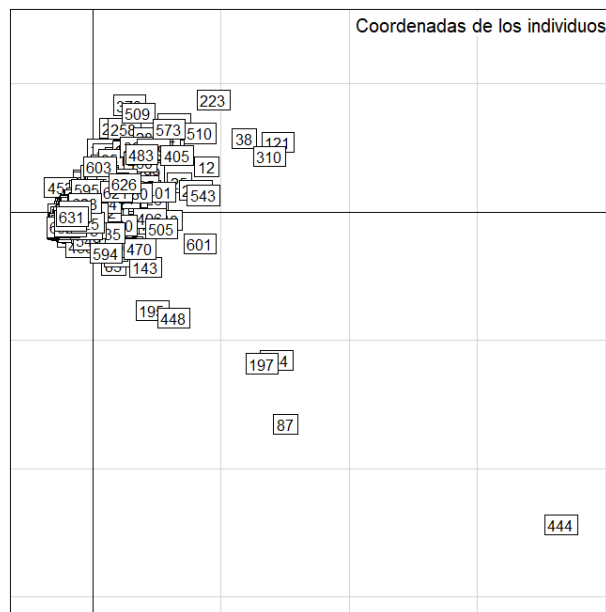


Figura 60. Coordenadas de los individuos respecto al CP1 y CP3. Adaptado de RStudio

CP2 y CP3

En la figura 61 se observa que máquinas como el pasteurizador, transportador de botellas e inspector de caja llena no son representados por estos dos componentes principales, dado que estos

son representados por el PC1. De acuerdo con la figura 61 se observa que el componente 3 representan las máquinas: envasadora e inspector de botellas vacías.

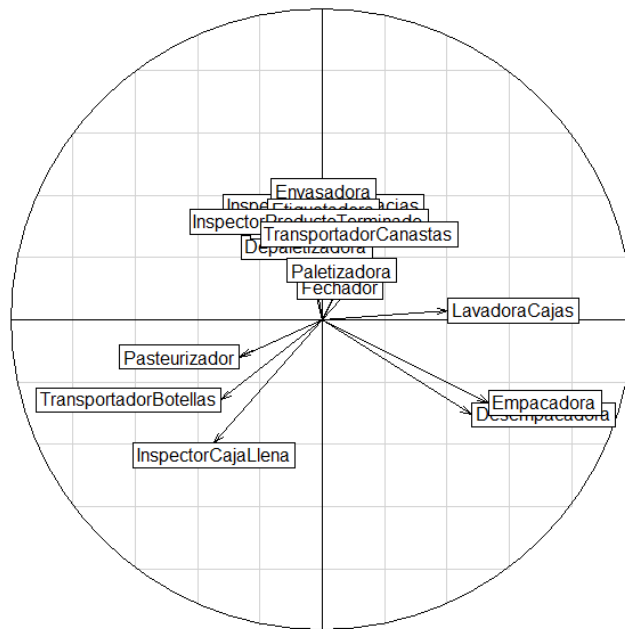


Figura 61. Ubicación de máquinas respecto a CP2 y CP3. Adaptado de RStudio

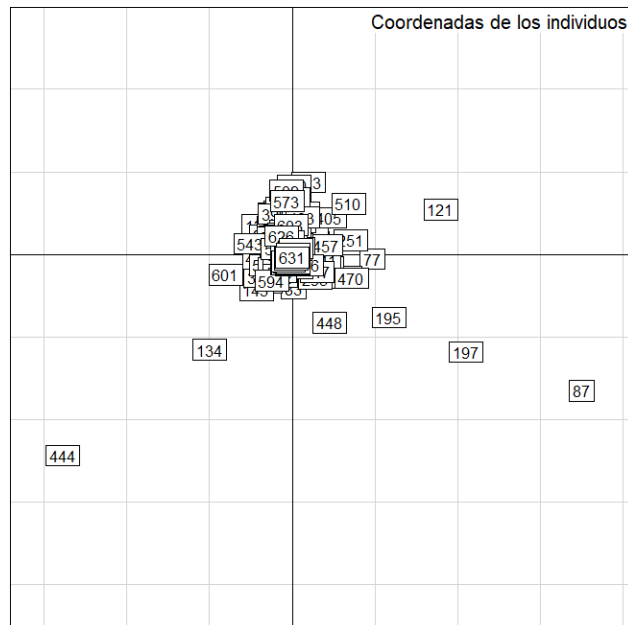


Figura 62. Coordenadas de los individuos respecto al CP1 y CP3. Adaptado de RStudio

7.3 Resultados regresión lineal múltiple

Realizado los grupos característicos y analizados con el programa SPSS se obtienen los siguientes resultados:

Coefficientes^a

Modelo	Coefficients no estandarizados		Coefficients estandarizados	t	Sig.	95,0% intervalo de confianza para B		Correlaciones			Estadísticas de colinealidad	
	B	Error estándar	Beta			Límite inferior	Límite superior	Orden cero	Parcial	Parte	Tolerancia	VIF
1 (Constante)	75,374	75,281		1,001	,319	-73,853	224,401					
Light-P1-Club	119,785	100,565	,168	1,191	,236	-79,293	318,863	-,004	,107	,087	,269	3,713
Light-P2	-66,547	40,415	-,149	-1,647	,102	-146,553	13,459	-,330	-,147	-,121	,656	1,525
Light-P2-Club	82,884	84,809	,150	,977	,330	-85,005	250,773	,036	,088	,072	,227	4,408
Light-P3-Club	150,403	89,614	,273	1,678	,096	-26,997	327,804	,257	,150	,123	,203	4,921
Pony-Light-P1-Club	225,808	135,285	,228	1,669	,098	-42,001	493,618	,257	,149	,122	,289	3,463
Pony-Light-P2	-74,453	83,462	-,116	-,892	,374	-239,674	90,769	-,033	-,081	-,065	,318	3,147
Pony-Light-P2-Club	65,807	116,451	,087	,565	,573	-164,720	296,334	,115	,051	,041	,228	4,390
Pony-Light-P3	-95,454	80,131	-,168	-1,191	,236	-254,081	63,173	-,080	-,107	-,087	,272	3,683
Pony-Light-P3-Club	45,527	108,585	,060	,419	,676	-169,428	260,481	,112	,038	,031	,262	3,817
Unidades Pony	,000	,000	,249	1,262	,209	,000	,000	,187	,114	,093	,138	7,222
Unidades Poker	-4,984E-7	,000	-,002	-,020	,984	,000	,000	-,085	-,002	-,001	,428	2,334
Unidades Pilsen	3,872E-5	,000	,133	1,204	,231	,000	,000	,181	,108	,088	,443	2,257
Unidades Aguila	-1,179E-5	,000	-,057	-,476	,635	,000	,000	-,014	-,043	-,035	,374	2,673
Unidades Club	-2,601E-5	,000	-,097	-,403	,688	,000	,000	,377	-,036	-,030	,093	10,724
Unidades Totales	3,075E-5	,000	,261	1,987	,049	,000	,000	,357	,177	,146	,311	3,216

a. Variable dependiente: Tiempo perdido

Figura 63. Coeficientes beta de las variables predictoras. Adaptado de SPSS Statistics 23

Se puede identificar (figura 63) que el grupo Pony-Light-P1-Club es el que representa mayor pérdida de tiempo con una beta de 225.8, seguido del grupo Light-P3-Club con una beta de 150.403 y el grupo Light-P1-Club con un valor beta de 119.785. Estos tres grupos son los más significativos en relación con el tiempo perdido durante el ciclo de producción. Los tres grupos representan algo en similar: presentan el producto Light y Club en sus grupos.

El programa arroja tres (3) modelos de predicción (ver figura 64).

Resumen del modelo^d

Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación	Estadísticos de cambio				
					Cambio en R cuadrado	Cambio en F	gl1	gl2	Sig. Cambio en F
1	,377 ^a	,142	,136	155,2333	,142	22,592	1	136	,000
2	,492 ^b	,242	,231	146,4751	,100	17,750	1	135	,000
3	,530 ^c	,281	,265	143,1938	,039	7,258	1	134	,008

a. Predictores: (Constante), Unidades Club

b. Predictores: (Constante), Unidades Club, Unidades Totales

c. Predictores: (Constante), Unidades Club, Unidades Totales, Pony-Light-P1-Club

d. Variable dependiente: Tiempo perdido

Figura 64. Resumen de modelo de regresión lineal múltiple. Adaptado de SPSS Statistics 23

De los tres modelos resultantes, se escoge el modelo 3 porque presenta el coeficiente de correlación mayor. Este R cuadrado nos indica que tan bien explica este modelo la variable dependiente, que es para este caso, el tiempo total perdido en cada ciclo de producción.

De acuerdo con la prueba ANOVA (ver figura 65), arroja para el modelo tres un valor de significancia menor a 0.05, el cual indica que este modelo predice de forma correcta la variable dependiente.

ANOVA^a

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	544413,937	1	544413,937	22,592	,000 ^b
	Residuo	3277244,207	136	24097,384		
	Total	3821658,145	137			
2	Regresión	925240,454	2	462620,227	21,562	,000 ^c
	Residuo	2896417,691	135	21454,946		
	Total	3821658,145	137			
3	Regresión	1074060,717	3	358020,239	17,461	,000 ^d
	Residuo	2747597,427	134	20504,458		
	Total	3821658,145	137			

a. Variable dependiente: Tiempo perdido

b. Predictores: (Constante), Unidades Club

c. Predictores: (Constante), Unidades Club, Unidades Totales

d. Predictores: (Constante), Unidades Club, Unidades Totales, Pony-Light-P1-Club

Figura 65. ANOVA de los modelos de regresión. Adaptado de SPSS Statistics 23

Coefficient

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error estándar	Beta		
1	(Constante)	278,150	16,555		16,802	,000
	Unidades Club	,000	,000	,377	4,753	,000
2	(Constante)	45,345	57,423		,790	,431
	Unidades Club	9,169E-5	,000	,341	4,515	,000
	Unidades Totales	3,743E-5	,000	,318	4,213	,000
3	(Constante)	34,690	56,276		,616	,539
	Unidades Club	7,837E-5	,000	,291	3,830	,000
	Unidades Totales	3,916E-5	,000	,333	4,497	,000
	Pony-Light-P1-Club	201,923	74,951	,204	2,694	,008

a. Variable dependiente: Tiempo perdido

Figura 66. Coeficientes de los modelos de regresión. Adaptado de SPSS Statistics 23

En la figura 66 se presentan los respectivos coeficientes para el modelo de predicción.

Diseñando el modelo matemático, da como resultado:

$$\text{Tiempo perdido} = 34,690 + (0,00007837 * \text{UnidadesClub}) + (0,00003916 * \text{UnidadesTotales}) + (201,293 * \text{PonyLightP1Club})$$

De la ecuación anterior, las variables significativas son:

Unidades club: Cantidad de unidades a elaborar del producto club colombia.

Unidades totales: unidades totales a producir durante el ciclo de producción.

PonyLightP1Club: Variable dummy que indica el orden de los productos en el ciclo de producción.

Cabe recordar que la variable PonyLightP1Club es categórica, por ende, tomara el valor de 1 el orden es el mismo y 0 si el orden es diferente al mostrado en el modelo.

Este modelo indica que el tiempo perdido en cada ciclo de producción depende de las unidades de club a producir, de las unidades totales y del grupo PonyLightP1Club que aporta 201,293 minutos si este grupo es elaborado.

Para el segundo modelo de predicción propuesto (ver figura 67 y figura 68) nos presenta un coeficiente de correlación de 0.303. El nuevo modelo presenta cinco (5) variables significativas (una numérica y cuatro dicotómicas). Se observa que tres de las variables categóricas presentan coeficiente negativo y una presenta coeficiente positivo, esto quiere decir que las variables que presentan valor negativo son recomendables elaborar en esa secuencia porque disminuyen el tiempo total perdido, mientras que, si presenta valor positivo, esta se refleja con un aumento en el tiempo total perdido al finalizar el ciclo de producción.

Resumen del modelo^f

Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
1	,359 ^a	,129	,122	156,3370
2	,450 ^b	,202	,191	150,1194
3	,501 ^c	,251	,235	145,9750
4	,528 ^d	,279	,257	143,8141
5	,550 ^e	,303	,277	141,9177

- a. Predictores: (Constante), Unidades Totales
- b. Predictores: (Constante), Unidades Totales, Pony-Light-P1-Club
- c. Predictores: (Constante), Unidades Totales, Pony-Light-P1-Club, Light-P2
- d. Predictores: (Constante), Unidades Totales, Pony-Light-P1-Club, Light-P2, Light-P3
- e. Predictores: (Constante), Unidades Totales, Pony-Light-P1-Club, Light-P2, Light-P3, Pony-Light-P3
- f. Variable dependiente: Tiempo total por grupo de productos

Figura 67. ANOVA del modelo de regresión 2. Adaptado de SPSS Statistics 23

Coefficientes^a

Modelo		Coefficients no estandarizados		Coefficients estandarizados	t	Sig.
		B	Error estándar	Beta		
1	(Constante)	57,990	61,204		,947	,345
	Unidades Totales	4,232E-5	,000	,359	4,495	,000
2	(Constante)	41,102	58,962		,697	,487
	Unidades Totales	4,375E-5	,000	,371	4,835	,000
	Pony-Light-P1-Club	270,446	76,240	,272	3,547	,001
3	(Constante)	105,343	61,274		1,719	,088
	Unidades Totales	3,641E-5	,000	,309	3,985	,000
	Pony-Light-P1-Club	250,108	74,450	,251	3,359	,001
	Light-P2	-103,151	34,709	-,231	-2,972	,004
4	(Constante)	111,286	60,425		1,842	,068
	Unidades Totales	3,889E-5	,000	,329	4,288	,000
	Pony-Light-P1-Club	229,347	73,923	,231	3,102	,002
	Light-P2	-122,735	35,280	-,274	-3,479	,001
	Light-P3	-65,216	28,914	-,175	-2,256	,026
5	(Constante)	126,923	60,071		2,113	,036
	Unidades Totales	3,900E-5	,000	,330	4,357	,000
	Pony-Light-P1-Club	213,050	73,343	,214	2,905	,004
	Light-P2	-138,980	35,629	-,311	-3,901	,000
	Light-P3	-81,603	29,537	-,219	-2,763	,007
	Pony-Light-P3	-93,035	43,353	-,163	-2,146	,034

a. Variable dependiente: Tiempo total por grupo de productos

Figura 68. Coeficientes para el modelo 2. Adaptado de SPSS Statistics 23

Una vez calculada la variable predictora, es necesario realizar una prueba T para muestras relacionadas (ver figura 69, figura 70 y figura 71), en este caso, entre el tiempo perdido real contra la nueva variable predicha.

Estadísticas de muestras emparejadas

		Media	N	Desviación estándar	Media de error estándar
Par 1	TimeLost	326,0602	139	88,41214	7,49902
	Tiempo perdido	326,590	139	166,8631	14,1531

Figura 69. Estadísticos para las muestras relacionadas: variable real vs variable nueva. Adaptado de SPSS Statistics 23

Correlaciones de muestras emparejadas

		N	Correlación	Sig.
Par 1	TimeLost & Tiempo perdido	139	,533	,000

Figura 70. Correlaciones de las muestras relacionadas: variable real vs variable nueva. Adaptado de SPSS Statistics 23

Prueba de muestras emparejadas

		Diferencias emparejadas				t	gl	Sig. (bilateral)	
		Media	Desviación estándar	Media de error estándar	95% de intervalo de confianza de la diferencia				
					Inferior				Superior
Par 1	TimeLost- Tiempo perdido	-,52971	141,21746	11,97791	-24,21367	23,15425	-,044	138	,965

Figura 71. Prueba T para las muestras relacionadas: variable real vs variable nueva. Adaptado de SPSS Statistics 23

La prueba T presenta medias similares entre las dos variables, presenta una correlación de 0.533, el cual es bueno. Finalmente, presenta un valor de significancia de 0.965 el indica que el modelo predice en un 96.5 % el valor de la variable dependiente real, es decir, el modelo es significativo y puede ser utilizado.

Una vez terminado el análisis, el modelo matemático definido es puesto en práctica con datos de año 2018. Se tomaron cuatro (4) semanas al azar. Los resultados son los siguientes:

Tabla 8.

Aplicación de modelo matemático para datos del año 2018

DATOS	TIEMPO REAL	TIEMPO PRONOSTICADO	% ERROR
Grupo: Poker-Light-Aguila-Club			
Unidades totales = 1.253.490	393 min	371,13 min	5,89%
Unidades club = 662.100			
Grupo: Light-Pilsen-Águila-Poker			
Unidades totales = 6.703.260	275 min	297,18 min	7,46%
Grupo: Pony-Light-Aguila-Pilsen-Poker-Club			
Unidades totales = 6.328.530	426 min	397,18	7,26%
Unidades club = 1.463.170			
Grupo: Light-Poker-Aguila-Pilsen-Club			
Unidades totales = 10.088.070	498 min	534,72	6,87%
Unidades club = 1339.570			

Con base a la tabla 8, se afirma la hipótesis de que el modelo matemático propuesto cumple a cabalidad su función, dado que presente errores pequeños, con un máximo de 7,46% de error.

8. Modelo estándar

Con base al paso a paso seguido para el desarrollo del proyecto se define un modelo estándar que puede ser aplicado en industrias de procesos similares. Este modelo consta de cuatro (4) etapas: recolección, ajuste, modelo y visualización y validación.

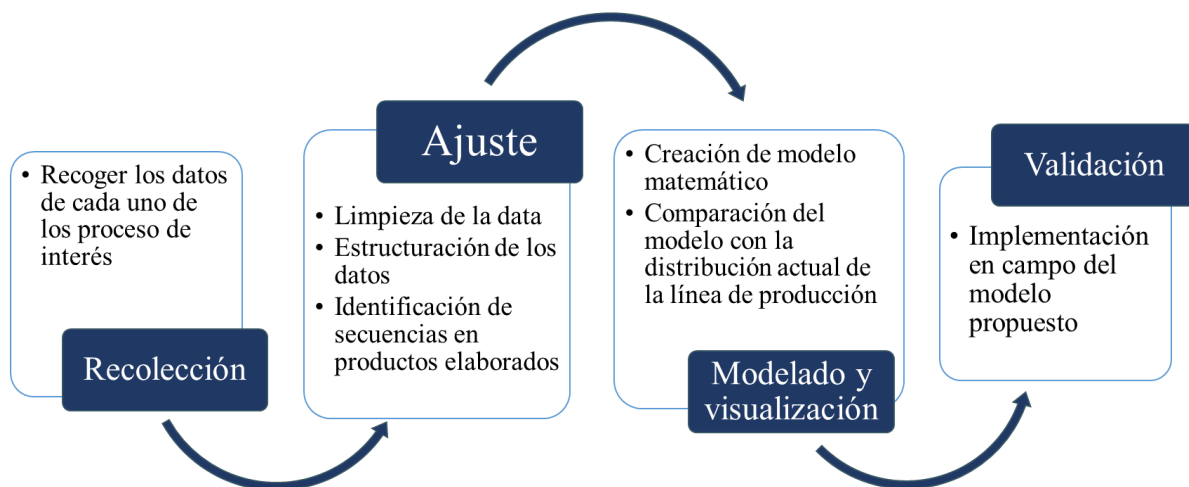


Figura 72. Modelo estándar

Cada etapa tiene una función importante en el proceso porque para pasar a la siguiente es indispensable que la etapa anterior esta cumplida correctamente. A continuación, se describen aspectos importantes para tener en cuenta para el cumplimiento de cada etapa.

Recolección: es la primera etapa y es la encargada de identificar los procesos de interés y recolectar los datos que serán utilizados para el desarrollo del proyecto.

Ajuste: la primera función de esta etapa es tener claro los resultados que se desean obtener, hecho esto, limpiar y estructurar los datos de tal forma que el software pueda procesarlos de forma correcta. En esta etapa, es necesario la creación, eliminación y modificación de variables a utilizar en cada técnica a utilizar. La parte final de la etapa consiste en la identificación de la secuencia de productos para la creación, si es necesario, de variables dummy.

Modelado y visualización: esta etapa es la encargada de la creación de los modelos matemáticos y la selección del que representa mejor la variable dependiente. Esta identificación se realiza mediante la comprensión de los valores arrojados y la comparación con los valores de la variable dependiente real. Adicional, se debe verificar si existe relación de los resultados con la distribución actual de las máquinas en la línea de producción y con esto, generar información para la creación de estrategias en las labores de mantenimiento.

Validación: para garantizar que el modelo propuesto representa de forma significativa la variable dependiente, este debe ser aplicado en campo con datos reales y contrarrestar el valor pronosticado con el valor real. De acuerdo con esto, tomar acciones de ajuste para el modelo.

9. Conclusiones

- A partir del conjunto de datos y el análisis básico realizado, se obtienen resultados naturales donde se observa que para ciertos productos dentro de la empresa se considera que los tiempos de perdida son muy parecidos. Esto se evidencia mediante la prueba estadística realizada para los productos Pilsen, Poker y Águila donde arroja como resultado que no hay diferencia entre medias.
- La técnica de componentes principales permite visualizar de forma sencilla y eficaz la relación existente entre las variables que comprenden cada factor y su ubicación real en la línea de producción. Este nuevo conocimiento brinda a la gerencia la oportunidad de generar nuevas estrategias para las labores de mantenimiento.
- Para la construcción del modelo de regresión se utilizaron como variables predictoras la secuencia de los productos como fueron elaborados y las unidades que se producen. Para

esto, se identificaron grupos de secuencias característicos donde permanecían fijos algunos productos, logrando así, disminuir el número de variables predictoras.

- El modelo de predicción propuesto resulta adecuado para predecir los tiempos de pérdida, donde se había considerado que el tiempo de pérdida dependía de la secuencia de los productos y la cantidad total de unidades elaboradas.

10. Recomendaciones

Abordando un área donde la minería de datos ha sido poco usada, se presentan factores reales que no se tienen en cuenta porque su utilización hace robusto el análisis. Dichas variables, tales como, mantenimiento programado de las diferentes máquinas, ciclo de vida de las máquinas, causas externas y/o ambientales; hacen posible una futura investigación mas a fondo sobre la problemática planteada.

Al igual que la minería de datos fue implementada en este proyecto para un proceso productivo continuo, se da un gran paso para iniciar futuras investigaciones donde existan procesos de producción similares tomando como base el análisis propuesto en este proyecto.

Se recomienda incentivar a los estudiantes para desarrollar problemas de investigación teórico prácticos donde se trabajen con datos reales brindando unión entre empresa, universidad y estudiante.

Referencias Bibliográficas

- Barón López, F. J., & Télez Montiel, F. (2004). Capítulo 6 : Regresión múltiple. *Apuntes de Bioestadística*, 35–43. Retrieved from <http://www.bioestadistica.uma.es/baron/apuntes/ficheros/cap06.pdf>
- Bavaria S.A. (2014). Curso de capacitación "RCM Bavaria". *Curso de capacitación "RCM Bavaria"(171)*. Bogotá, Colombia.
- Bavaria S.A. (2016). *Portal Corporativo Bavaria S.A.* Obtenido de <http://www.bavaria.co/acerca-de-nosotros/historia-bavaria>
- Choudhary, A. K., Harding, J. A., & Tiwari, M. K. (2009). Data mining in manufacturing: A review based on the kind of knowledge. *Journal of Intelligent Manufacturing*, 20(5), 501–521. <https://doi.org/10.1007/s10845-008-0145-x>
- Dinero, R. (2017). Así se ve el panorama de Colombia en 2018. *14-12-2017*.
- Dorian Pyle. (1999). *Data preparation for data mining*. Morgan Kaufmann Publishers.
- Eduardo, P. (1999). Análisis de Componentes Principales. *Proyecto E-Math Financiado Por La Secretaría de Estado de Educación Y Universidades (MECD)*, 141–151. <https://doi.org/10.1016/j.infsof.2008.09.005>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), 27–34. <https://doi.org/10.1145/240455.240464>
- García, H. J., & López, J. M. M. (2012). Técnicas De Análisis De Datos Aplicaciones Prácticas Utilizando Microsoft Excel Y Weka, 5–43. Retrieved from <http://ocw.uc3m.es/ingenieria->

informatica/analisis-de-datos/libroDataMiningv5.pdf

Gómez Torrejano, M. (2009). Aplicación de la Minería de Datos para describir las ocurrencias de fallas en los equipos y sistemas de la plataforma tecnológica de PDVSA AIT Servicios Comunes Centro. *Univ de Los Andes Merida*, 33–47 (DM–2009–1). Retrieved from http://webdelprofesor.ula.ve/economia/gcolmen/programa/economia/work_paper_sistemas_manuel_gomez.pdf

Han, J., Kamber, M., & Pei, J. (Computer scientist). (2012). *Data mining: concepts and techniques*. Elsevier/Morgan Kaufmann. <https://doi.org/10.1017/CBO9781107415324.004>

Harding. (2001). Data mining techniques for supporting manufacturing enterprise design. *Internacional conference on industrial and production management*. Quebec.

Isabel Alonso, M. L., & Ibañez, D. (2017). Información Industria Septiembre 2017 Oficina de Estudios Económicos.

Kang, S., Kim, E., Shim, J., Cho, S., Chang, W., & Kim, J. (2017). Mining the relationship between production and customer service data for failure analysis of industrial products. *Computers and Industrial Engineering*, 106, 137–146. <https://doi.org/10.1016/j.cie.2017.01.028>

Kopanakis, I., & Theodoulidis, B. (2003). Visual data mining modeling techniques for the visualization of mining outcomes. *Journal of Visual Languages and Computing*, 14(6), 543–589. <https://doi.org/10.1016/j.jvlc.2003.06.002>

Larose, D. T. (2004). *Discovering knowledge in data: an introduction to data mining*.

Mosaddar, D., & Shojaie, A. A. (2013). A data mining model to identify inefficient maintenance activities. *International Journal of Systems Assurance Engineering and Management*, 4(2), 182–192. <https://doi.org/10.1007/s13198-013-0148-7>

Ortiz, E. (2015). Regresión Lineal Múltiple: Introducción. Retrieved from

http://humanidades.cchs.csic.es/cchs/web_UAE/tutoriales/PDF/Regresion_lineal_multiple_3.pdf

Paolo, G. (2003). *Applied data mining: statistical methods for business and industry*. J. Wiley.

Pla, L. E. (1986). *Análisis multivariado: Método de componentes principales*. Coro, Falcon, Venezuela.

Vélez Torres, A. C., & Hernández R, J. A. (2001). *UN MODELO GENÉTICO PARA MINERÍA DE DATOS* Por: Ana Clara Vélez Torres Director: Prof. Jesús Antonio Hernández R. Tesis de Gr. Universidad Nacional de Colombia.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann. <https://doi.org/10.1016/C2009-0-19715-5>