



DESCRIPCIÓN DE DATOS

Análisis Comparativo de Servicios y Facturación entre Empresas

Autor:

Sergio Alejandro Romero López

Índice

| | |
|---|----|
| 1. Planteamiento del problema | 6 |
| 2. Descripción del conjunto de datos | 6 |
| 3. Análisis descriptivo de los datos..... | 7 |
| 3.1. Análisis Descriptivo Univariante..... | 7 |
| 3.1.1. Variable cualitativa | 7 |
| 3.1.2. Variables cuantitativas..... | 7 |
| 3.2. Análisis de la Variable Tiempo..... | 12 |
| 3.2.1. Análisis de variable Tiempo y Tipo de Empresa..... | 12 |
| 3.2.2. Análisis de variable Tiempo y Número de Servicios | 14 |
| 3.3. Análisis de la Variable Facturado..... | 16 |
| 3.3.1. Análisis de Variable Facturado y Tipo de Empresa | 16 |
| 3.3.2. Análisis de Variable Facturado y Número de Servicios | 18 |
| 3.4. Relación Bivariante entre Tiempo y Facturado | 20 |
| 3.5. Análisis con Distinción de Tipo de empresa | 21 |
| 4. Conclusiones | 22 |
| 5. Bibliografía | 23 |
| 6. Anexos | 24 |

Lista de Figuras

Pág.

| | |
|---|----|
| Figura 1. Distribución de Tipos de Empresa (Gráfico de Barras y Circular)..... | 7 |
| Figura 2. Curva de Densidad del Número de Servicios | 8 |
| Figura 3. Histograma de Número de Servicios (Frecuencias Absolutas y Relativas) | 8 |
| Figura 4. Curva de Densidad del Tiempo..... | 9 |
| Figura 5. Histograma del Tiempo (Frecuencias Absolutas y Relativas) | 9 |
| Figura 6. Curva de Densidad del Monto Facturado | 10 |
| Figura 7. Histograma del Monto Facturado (Frecuencias Absolutas) | 10 |
| Figura 8. Boxplot del Tiempo según Tipo de Empresa | 12 |
| Figura 9. Media del Tiempo por Tipo de Empresa (Gráfico de Barras) | 12 |
| Figura 10. Diagrama de Dispersión: Tiempo vs Tipo de Empresa..... | 13 |
| Figura 11. Boxplot del Tiempo según Número de Servicios | 14 |
| Figura 12. Tiempo Promedio por Número de Servicios Contratados (Gráfico de Barras) | 14 |
| Figura 13. Diagrama de Dispersión: Tiempo vs Tipo de Empresa..... | 15 |
| Figura 14. Boxplot del Facturado según Tipo de Empresa | 16 |
| Figura 15. Promedio de Facturado según Tipo de Empresa | 16 |
| Figura 16. Diagrama de Dispersión: Facturado vs Tipo de Empresa | 17 |
| Figura 17. Boxplot del Facturado según Número de Servicios | 18 |
| Figura 18. Promedio de Facturado según Número de Servicios | 18 |

| | |
|---|----|
| Figura 19. Diagrama de Dispersión: Facturado vs Número de Servicios | 18 |
| Figura 20. Relación Positiva entre Facturado y Tiempo (Diagrama de Dispersión) | 20 |
| Figura 21. Modelo de Regresión Lineal: Facturado vs Tiempo | 20 |
| Figura 22. Medidas Estadísticas del Tiempo según Tipo de Empresa | 21 |
| Figura 23. Medidas Estadísticas de Facturado según Tipo de Empresa | 21 |

Lista de Tablas

| | Pág. |
|--|------|
| Tabla 1. Distribución de Tipo de Empresa | 7 |
| Tabla 2. Distribución de Número de Servicios..... | 7 |
| Tabla 3. Medidas de Dispersión para el Número de Servicios | 8 |
| Tabla 4. Distribución Continua del Tiempo | 9 |
| Tabla 5. Medidas de Dispersión del Tiempo | 9 |
| Tabla 6. Distribución Continua del Monto Facturado | 10 |
| Tabla 7. Medidas de Dispersión de Facturado | 10 |

Lista de Anexos

| | Pág. |
|--|-------------|
| Anexo A. Para cargar datos y mostrarlos por variable | 24 |
| Anexo B. Explorar datos | 24 |
| Anexo C. Funciones..... | 24 |
| Anexo D. Análisis Descriptivo Univariante..... | 25 |
| Anexo E. Análisis de la Variables Tiempo | 25 |
| Anexo F. Análisis de la Variable Facturado..... | 26 |
| Anexo G. Correlación entre el facturado, tipo de servicio, número de servicio..... | 26 |
| Anexo H. Relación Bivariante entre Tiempo y Facturado | 26 |
| Anexo I. Análisis con Distinción de Tipo de empresa | 26 |

1. Planteamiento del problema

Una empresa de telecomunicaciones desea analizar su base de clientes para comprender mejor su perfil. Para ello, se han recopilado datos de una muestra de empresas clientes que incluyen información sobre las siguientes variables:

- *Tipo de empresa (pyme o empresa grande).*
- *Número de servicios contratados (teléfono fijo, móvil, internet, datos, tv, etc).*
- *Tiempo de permanencia (en meses).*
- *Total facturado (en euros).*

El objetivo de la empresa es realizar un análisis descriptivo de estas variables para identificar patrones y tendencias en la base de clientes de la empresa y proporcionar información valiosa para futuras estrategias de marketing y retención de clientes.

2. Descripción del conjunto de datos

Tras el análisis y observaciones correspondientes, se puede observar que el conjunto de datos se compone de 207 registros y 4 variables: Tipo de Empresa, Número de Servicios, Tiempo, y Facturado. Cada una de estas variables tiene un tipo de dato específico según su naturaleza.

La variable **Tipo de Empresa** es de tipo carácter y es **cualitativa de tipo nominal o categórica**, ya que no sigue un orden jerárquico y se divide en dos categorías: "pyme" y "grande", mientras que **Número de Servicios** es de tipo entero y es **cuantitativa discreta**, dado que representa un conteo y no acepta valores decimales. Las variables **Tiempo** y **Facturado** son numéricas y son **cuantitativas continuas**, ya que pueden tomar valores decimales y representar mediciones exactas. Por lo tanto, ambas reflejan valores numéricos con mayor precisión.

Esta clasificación se ha definido en función de la estructura y características observadas durante la recolección y organización de los datos.

3. Análisis descriptivo de los datos

3.1. Análisis Descriptivo Univariante

Luego de llevar a cabo el estudio del tipo de variable y su distribución, se procedió a realizar las tablas de frecuencias y distribuciones pertinentes utilizando **RStudio** como herramienta de análisis estadístico.

A partir de esto, se demuestran el comportamiento de las variables estudiadas, ofreciendo una base sólida para su interpretación.

3.1.1. Variable cualitativa

| Tipo de empresa | Frecuencia abs (ni) | Frecuencia rel (fi) |
|-----------------|---------------------|---------------------|
| grande | 75 | 0.36 |
| pyme | 132 | 0.64 |
| TOTAL | 207 | 1 |

Tabla 1. Distribución de Tipo de Empresa

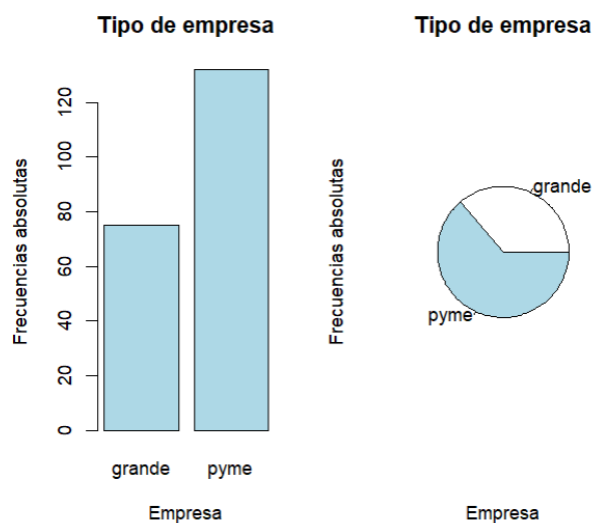


Figura 1. Distribución de Tipos de Empresa (Gráfico de Barras y Circular)

3.1.2. Variables cuantitativas

Número de servicios

| Número de servicios | Frecuencia abs (ni) | Frecuencia abs AC (Ni) | Frecuencia rel (fi) | Frecuencia rel AC (Fi) |
|---------------------|---------------------|------------------------|---------------------|------------------------|
| 2 | 23 | 23 | 0.11 | 0.11 |
| 3 | 39 | 62 | 0.19 | 0.3 |
| 4 | 39 | 101 | 0.19 | 0.49 |
| 5 | 44 | 145 | 0.21 | 0.7 |
| 6 | 32 | 177 | 0.15 | 0.86 |
| 7 | 30 | 207 | 0.14 | 1 |
| TOTAL | 207 | 207 | 1 | 1 |

Tabla 2. Distribución de Número de Servicios

| Rango | Mediana | Media | Varianza | Desviación estadar | Coficiente de variación |
|-------|---------|-------|----------|--------------------|-------------------------|
| 5 | 5 | 4.55 | 2.48 | 1.58 | 34.73 |

Tabla 3. Medidas de Dispersión para el Número de Servicios

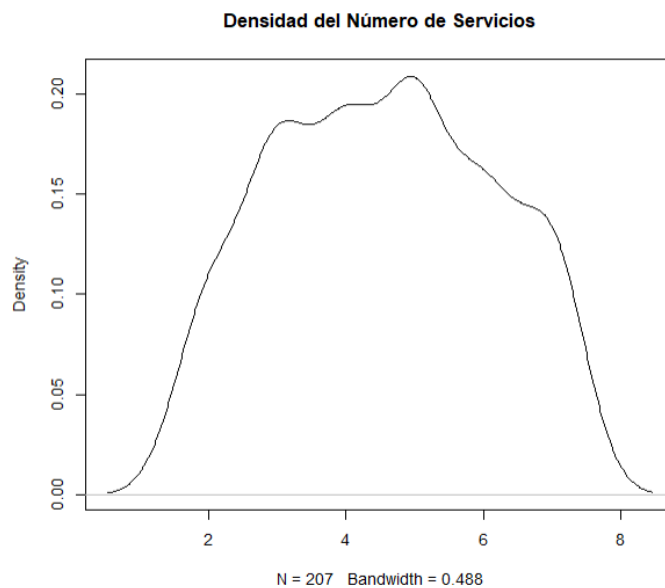


Figura 2. Curva de Densidad del Número de Servicios

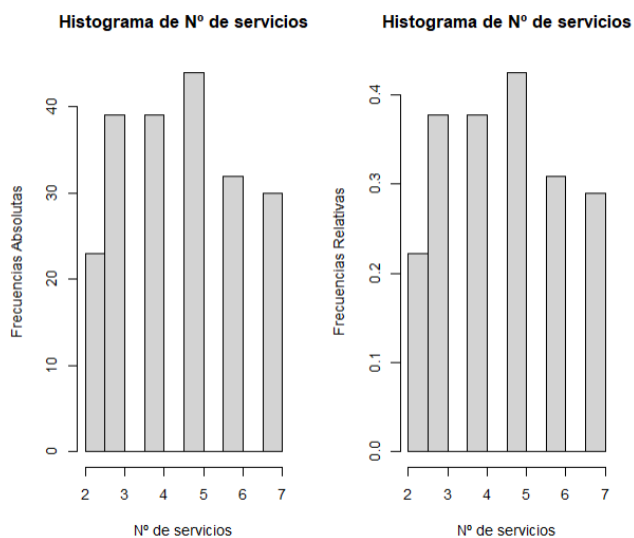


Figura 3. Histograma de Número de Servicios (Frecuencias Absolutas y Relativas)

Tiempo

| Límites | Marca de clase (Xi) | Frecuencia abs (ni) | Frecuencia abs AC (Ni) | Frecuencia rel (fi) | Frecuencia rel AC (Fi) |
|-----------------|---------------------|---------------------|------------------------|---------------------|------------------------|
| [60.23,91.12] | 75.67 | 12 | 12 | 0.06 | 0.06 |
| (91.12,122.01] | 106.56 | 25 | 37 | 0.12 | 0.18 |
| (122.01,152.9] | 137.45 | 33 | 70 | 0.16 | 0.34 |
| (152.9,183.79] | 168.34 | 48 | 118 | 0.23 | 0.57 |
| (183.79,214.68] | 199.23 | 38 | 156 | 0.18 | 0.76 |
| (214.68,245.57] | 230.12 | 32 | 188 | 0.16 | 0.91 |
| (245.57,276.46] | 261.02 | 16 | 204 | 0.08 | 0.99 |
| (276.46,307.35] | 291.91 | 2 | 206 | 0.01 | 1 |
| (307.35,338.24] | 322.8 | 0 | 206 | 0 | 1 |
| TOTAL | | 206 | 206 | 1 | 1 |

Tabla 4. Distribución Continua del Tiempo

| Rango | Mediana | Media | Varianza | Desviación estadar | Coefficiente de variación |
|--------|---------|--------|----------|--------------------|---------------------------|
| 278.02 | 171.89 | 172.72 | 2850.44 | 53.39 | 30.91 |

Tabla 5. Medidas de Dispersión del Tiempo

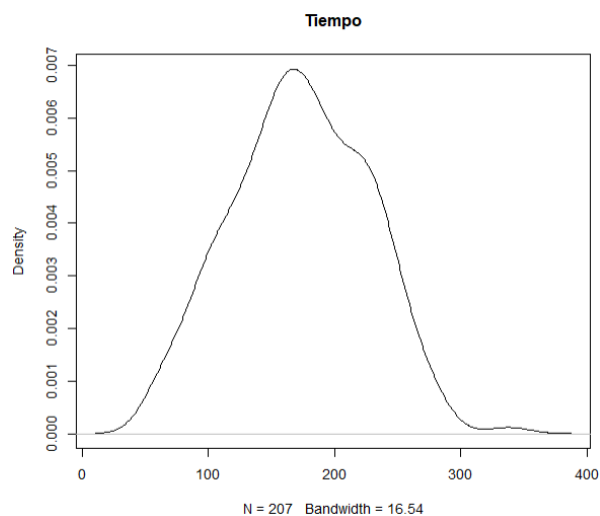


Figura 4. Curva de Densidad del Tiempo

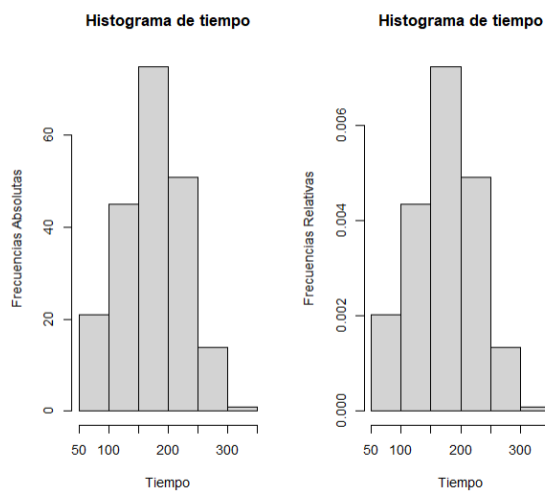


Figura 5. Histograma del Tiempo (Frecuencias Absolutas y Relativas)

Facturado

| Limites | Marca de clase (Xi) | Frecuencia abs (ni) | Frecuencia abs AC (Ni) | Frecuencia rel (fi) | Frecuencia rel AC (Fi) |
|-----------------|---------------------|---------------------|------------------------|---------------------|------------------------|
| [3816.5,6587.4] | 5201.95 | 14 | 14 | 0.07 | 0.07 |
| (6587.4,9358.2] | 7972.79 | 16 | 30 | 0.08 | 0.15 |
| (9358.2,12129] | 10743.63 | 26 | 56 | 0.13 | 0.27 |
| (12129,14900] | 13514.47 | 43 | 99 | 0.21 | 0.48 |
| (14900,17671] | 16285.31 | 38 | 137 | 0.18 | 0.67 |
| (17671,20442] | 19056.15 | 40 | 177 | 0.19 | 0.86 |
| (20442,23212] | 21826.99 | 14 | 191 | 0.07 | 0.93 |
| (23212,25983] | 24597.83 | 6 | 197 | 0.03 | 0.96 |
| (25983,28754] | 27368.67 | 9 | 206 | 0.04 | 1 |
| TOTAL | | 206 | 206 | 1 | 1 |

Tabla 6. Distribución Continua del Monto Facturado

| Rango | Mediana | Media | Varianza | Desviación estadar | Coficiente de variación |
|----------|----------|----------|-------------|--------------------|-------------------------|
| 24937.57 | 15213.88 | 15262.04 | 29851427.04 | 5463.65 | 35.8 |

Tabla 7. Medidas de Dispersión de Facturado

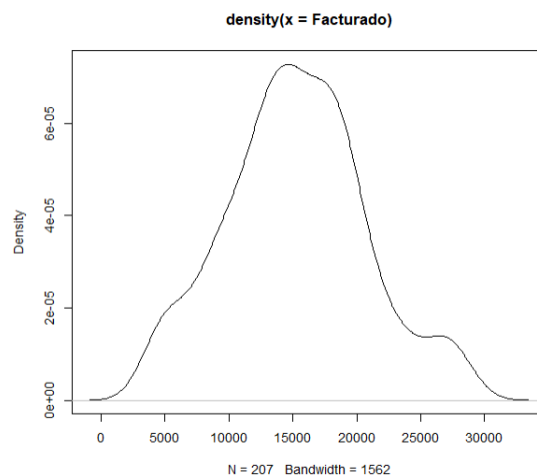


Figura 6. Curva de Densidad del Monto Facturado

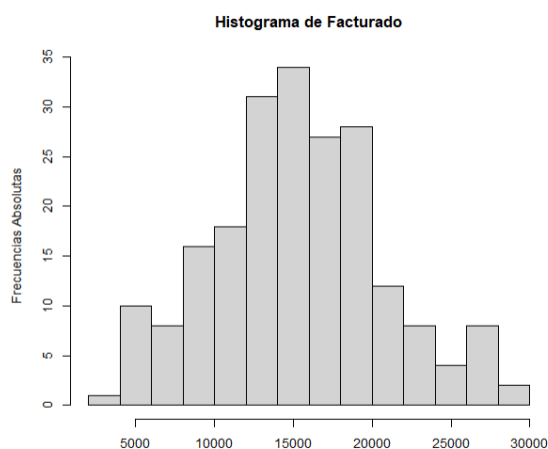


Figura 7. Histograma del Monto Facturado (Frecuencias Absolutas)

Interpretación:

- Tras crear la tabla de frecuencias en **R**, se identifican dos categorías: **empresas grandes** y **empresas pymes**. Las empresas grandes tienen una frecuencia absoluta de **75**, lo que representa aproximadamente el **36.2%** del total, mientras que las pymes presentan una frecuencia absoluta de **132**, lo que equivale al **63.8%**. Esto indica que una porción significativa de las empresas analizadas son pymes, mientras que un menor porcentaje corresponde a empresas grandes.
- Se observa que el coeficiente de variación es bajo, lo que indica una menor dispersión o variabilidad en relación con la media. Al comparar los tres valores de dispersión, la relación se establece de la siguiente manera: **CV (facturas) > CV (N°Serv) > CV(Tiempo)**.
- Tras analizar la gráfica de densidad y los cálculos de asimetría en R, se concluye que la distribución del número de servicios es casi simétrica perfecta o asimetría nula, con media, mediana y moda muy cercanas. Lo mismo ocurre en las variables de tiempo y facturación.

3.2. Análisis de la Variable Tiempo

A continuación, describiremos gráficamente la variable Tiempo, realizando un análisis los distintos tipos de variable.

3.2.1. Análisis de variable Tiempo y Tipo de Empresa

Describimos gráficamente la variable Tiempo, usando R, vamos a distinguir entre "Tipo de empresa" y "Número de Servicios".

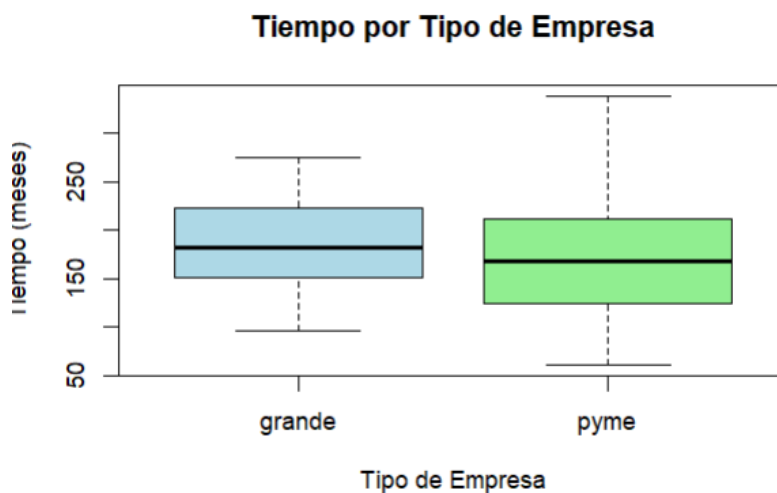


Figura 8. Boxplot del Tiempo según Tipo de Empresa

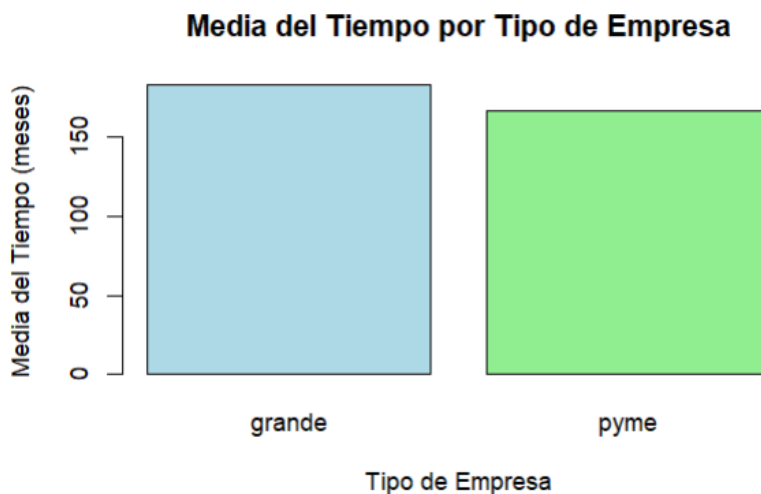


Figura 9. Media del Tiempo por Tipo de Empresa (Gráfico de Barras)

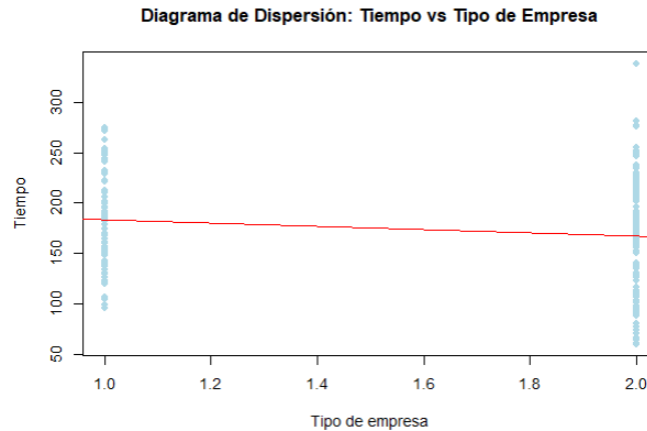


Figura 10. Diagrama de Dispersión: Tiempo vs Tipo de Empresa

Interpretación:

- Se puede deducir a simple vista, que la mediana del tiempo respecto a empresas grandes es por poco superior a la de las pymes, también podemos ver que la dispersión(variabilidad) de tiempo en las pymes es superior que en las empresas grandes y que no hay valores atípicos en la gráfica.
- Podemos observar en el grafico de barras que las empresas grandes tienen un promedio superior respecto a pymes, esto indica que tienen un periodo de promedio más largo.
- Podemos establecer “Tiempo” como la variable dependiente y “Tipo de Empresa” como la independiente y observamos que no hay algún tipo de relación lineal, y calculando su correlación entre las 2 variables nos resulta un numero **-0.1283312** cercano a 0, de ahí la gráfica casi como una lineal horizontal, ósea que **tiempo no depende** prácticamente de tipo de empresa.

3.2.2. Análisis de variable Tiempo y Número de Servicios

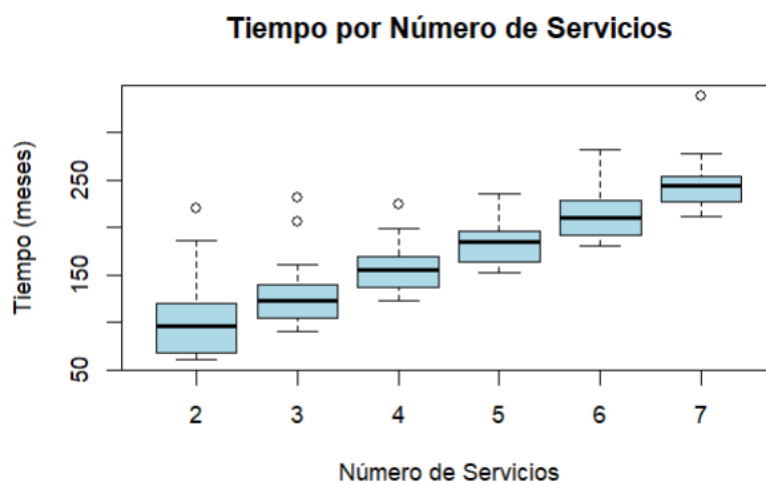


Figura 11. Boxplot del Tiempo según Número de Servicios

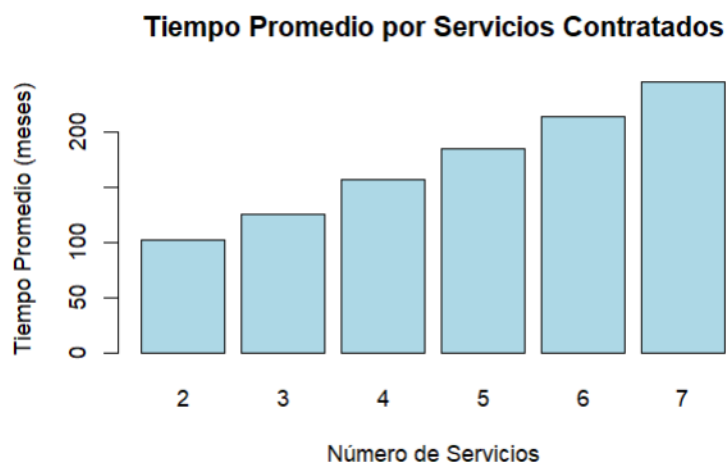


Figura 12. Tiempo Promedio por Número de Servicios Contratados (Gráfico de Barras)

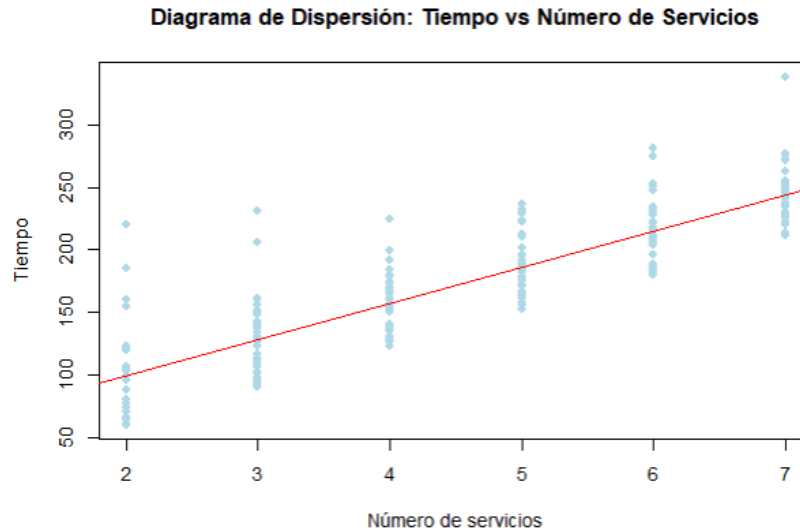


Figura 13. Diagrama de Dispersión: Tiempo vs Tipo de Empresa

Interpretación:

- Se puede deducir a simple vista, que la mediana del tiempo respecto a números de servicios aumenta, cuando el tiempo tiende a incrementarse, se observa también que hay valores atípicos, resaltados en servicio de 2,3,4,7 y como ultima conclusión que mientras parecía que tendiera la relación a una recta (correlación positiva > 0). dispersión(variabilidad) de tiempo en las pymes es superior que en las empresas grandes y que no hay valores atípicos en la gráfica.
- Podemos observar en el gráfico de barras que las empresas grandes tienen un promedio superior respecto a pymes, esto indica que tienen un periodo de promedio más largo, respecto al tiempo.
- En la segunda grafica podemos deducir que a través y más servicios contratados el promedio aumenta, teniendo una subida y también que a más servicios hay más tiempo de permanencia.
- Se observa una relación más lineal deduciendo que los clientes que contratan más servicios tienden a quedarse más tiempo con la empresa. La correlación entre ambas es 0.8605349 por lo que las variables Tiempo y Servicios **son dependientes**.

3.3. Análisis de la Variable Facturado

Como hemos interpretado anteriormente, procederemos a hacerlo para la siguiente variable y con sus graficas correspondientes.

3.3.1. Análisis de Variable Facturado y Tipo de Empresa

Describimos gráficamente la variable Facturado, usando R, vamos a distinguir entre "Tipo de empresa" y "Número de Servicios".

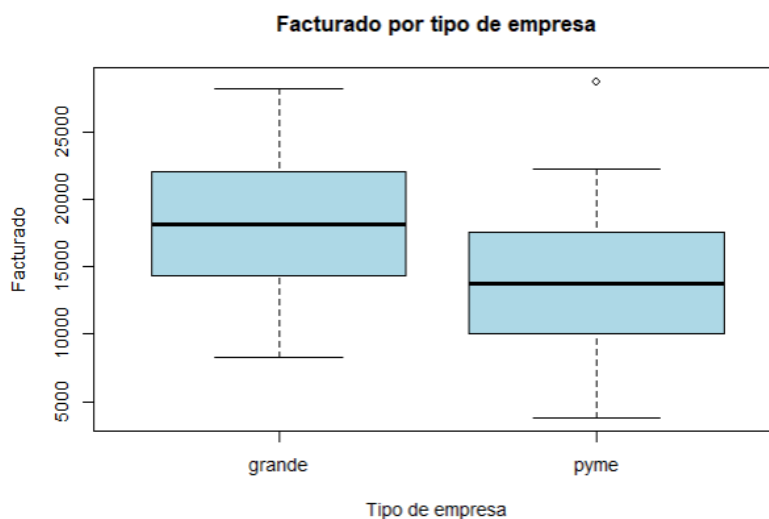


Figura 14. Boxplot del Facturado según Tipo de Empresa

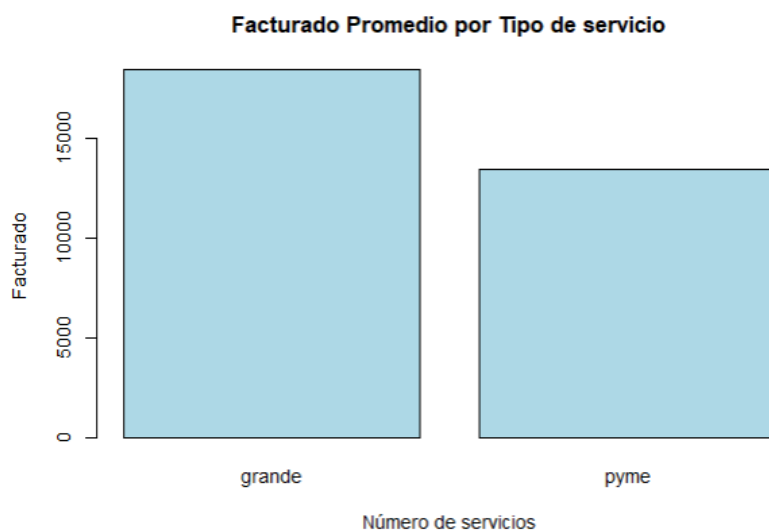


Figura 15. Promedio de Facturado según Tipo de Empresa

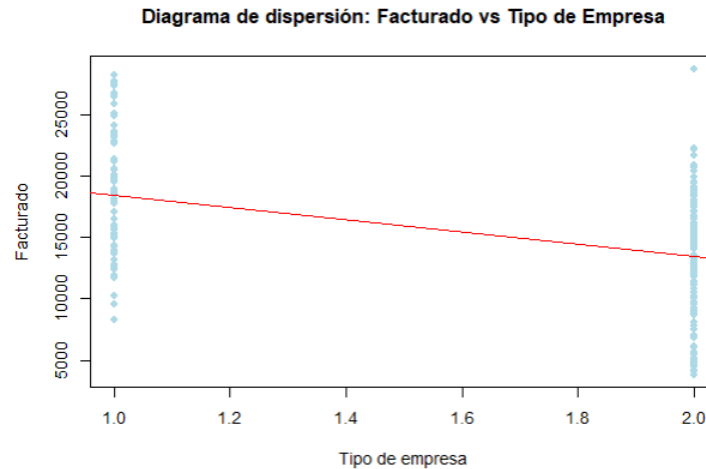


Figura 16. Diagrama de Dispersión: Facturado vs Tipo de Empresa

Interpretación:

- Se puede observar que la mediana de facturación de las empresas grandes es mayor que de las pymes y que aproximadamente ella ronda es menor que 20000 y mayor que 15000, y que para pymes ronda entre los 15000.
- Se puede observar también que las grandes empresas tienen una mayor variabilidad y que las pymes presentan valores atípicos, lo que indica que hay un facturado mas mayor que al resto.
- El valor facturado **tiene una dependencia moderada** del tipo de empresa, ya que el coeficiente de correlación de **-0.4100544** indica una relación negativa. Esto quiere decir que, aunque hay una tendencia general a que el valor facturado disminuya con el tipo de empresa, la relación no es lo suficientemente fuerte como para considerarla un factor determinante.

3.3.2. Análisis de Variable Facturado y Número de Servicios

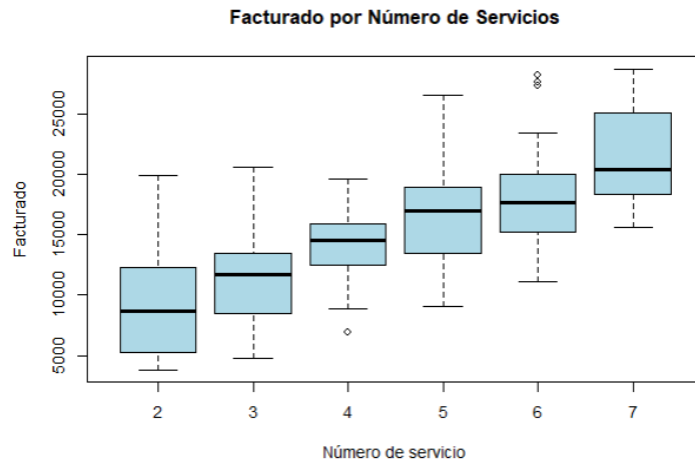


Figura 17. Boxplot del Facturado según Número de Servicios

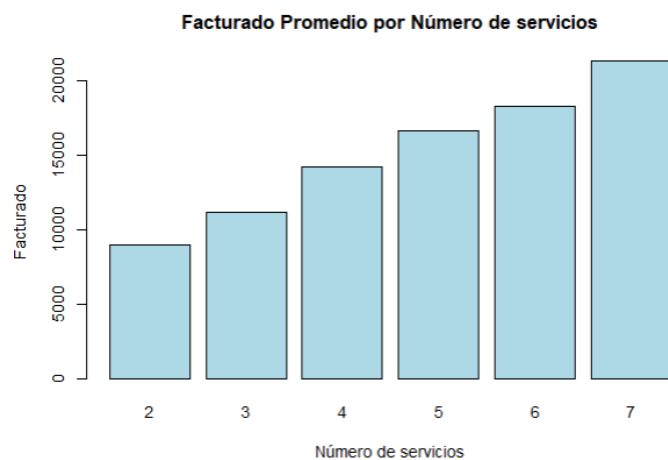


Figura 18. Promedio de Facturado según Número de Servicios

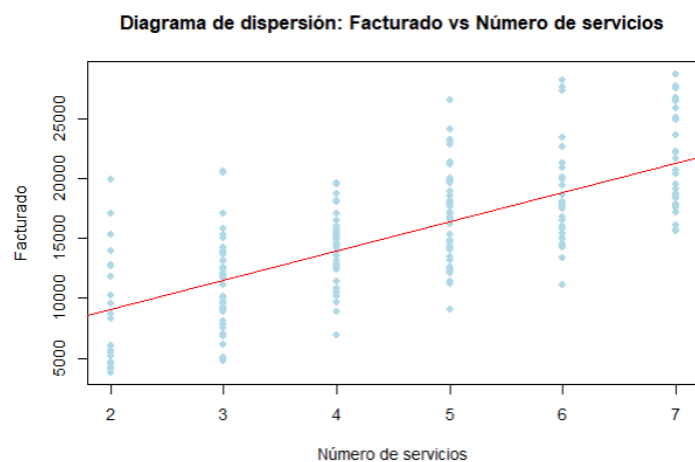


Figura 19. Diagrama de Dispersión: Facturado vs Número de Servicios

Interpretación:

- Se puede observar que a medida que aumenta los servicios, el total del facturado tiende a aumentar, entonces se deduce que a mas servicios genera mayor volumen de facturación.
- Se puede ver también algunos valores atípicos, fuera del rango establecido (Los bigotes). Estos puntos representan usuarios cuyo monto facturado es notablemente mayor o menor en comparación con el rango típico para ese número de servicios.
- El valor facturado **depende en parte del número de servicios**, ya que el coeficiente de correlación calculado es de **0.7041209**, lo que indica una relación positiva fuerte entre ambas variables, esto quiere decir que a medida que aumente el número de servicios, el valor facturado también tiende a aumentar.

3.4. Relación Bivalente entre Tiempo y Facturado

A continuación, realizaremos un estudio de relación bivalente entre Tiempo y Facturado con lo cual utilizaremos los diferentes tipos de gráficos para ver su relación.

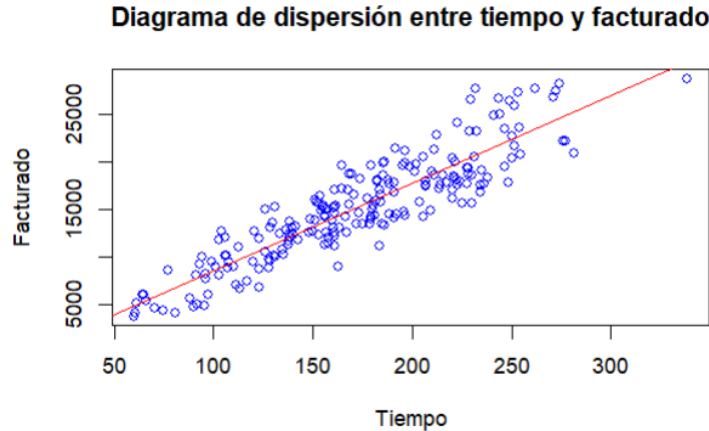


Figura 20. Relación Positiva entre Facturado y Tiempo (Diagrama de Dispersión)

```
> reg.Fact.Tiem<-lm(Facturado~Tiempo,data = datos)
> summary(reg.Fact.Tiem)

Call:
lm(formula = Facturado ~ Tiempo, data = datos)

Residuals:
    Min       1Q   Median       3Q      Max
-5281  -1798   -141    1723    6925

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -645.651    563.074   -1.147   0.253
Tiempo         92.103      3.115   29.564 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2387 on 205 degrees of freedom
Multiple R-squared:  0.81,    Adjusted R-squared:  0.8091
F-statistic: 874 on 1 and 205 DF, p-value: < 2.2e-16
```

Figura 21. Modelo de Regresión Lineal: Facturado vs Tiempo

Interpretación:

- Se puede observar que **a medida que aumenta la variable facturada, el tiempo tiende a aumentar**, entonces se deduce que a más tiempo genera mayor volumen de facturación, con esto no quiere decir que existe una tendencia de alza, ya que no todos los puntos se alinean perfectamente, también se observa que la variable facturada es la dependiente y tiempo la independiente.
- Se puede ver también que va entre un rango de 50-300 meses de facturado y que existen **pocos valores atípicos**.
- Hay calcular su correlación ya sea, por R-STUDIO o echo a mano, se puede ver que el resultado nos bota un numero de **0.8987834** que indicia una relación fuerte y positiva entre ambas variables, casi llegando al 1 (**Correlación positiva**). En la última imagen podemos ver la relación de la recta.

3.5. Análisis con Distinción de Tipo de empresa

Para analizar la relación entre el tiempo y lo facturado distinguiendo entre los tipos de empresas. A continuación, presentaremos una imagen, con los posibles resultados de las variables (Tiempo, Facturado) respecto al tipo de empresa.

| Tipo.de.empresa | Media_Tiempo | Mediana_Tiempo | Varianza_Tiempo | Desviacion_Tiempo |
|-----------------|--------------|----------------|-----------------|-------------------|
| 1 grande | 183. | 182. | 2156. | 46.4 |
| 2 pyme | 167. | 168. | 3169. | 56.3 |

Figura 22. Medidas Estadísticas del Tiempo según Tipo de Empresa

| Tipo.de.empresa | Media_Facturado | Mediana_Facturado | Varianza_Facturado | Desviacion_Facturado |
|-----------------|-----------------|-------------------|--------------------|----------------------|
| 1 grande | 18417.25 | 18140.87 | 26316823 | 5129.992 |
| 2 pyme | 13469.31 | 13797.13 | 23137878 | 4810.185 |

Figura 23. Medidas Estadísticas de Facturado según Tipo de Empresa

Interpretación:

- Se puede ver en la primera imagen, que las empresas grandes tienden a tener un tiempo promedio mayor que la de una pyme ($183 > 167$) también que la varianza y desviación estándar son mayores en la pyme, lo que nos dice que existe mayor dispersión de tiempo, en comparación de las empresas grandes.
- En la segunda imagen, se observa una mayor facturación en las pymes en la media y mediana con respecto a las empresas grandes. También que hay mayor dispersión en las empresas grandes por su mayor varianza y desviación.
- Donde se dice que las empresas grandes, que tienen mayor tiempo de media, tienden a facturar más. En contraste las pymes tienden a tener servicios más bajos y también facturan menos. La relación entre ambas **puede estar influenciada por el tipo de empresa**, ya que las empresas grandes muestran mayores valores en ambas variables. En resumen, **el tipo de empresa influye** en la relación de Tiempo y Facturado, siendo las empresas grandes la que invierten más tiempo y obtienen una mayor facturación, mientras que las pymes se muestran con más variabilidad en relación con los tiempos y facturan menos en promedio.

4. Conclusiones

A modo de cierre de este trabajo, podemos señalar que el análisis de los datos nos explica que las empresas pymes son las que dominan el total, debido a su gran presencia en los números de servicios que contratan. Cabe resaltar también que las empresas grandes tienden a tener un tiempo de permanencia más largo que las pymes, lo que señala mejor retención.

El resultado más relevante de esta conclusión es la correlación entre número de servicios y tiempo de permanencia, que sugiere que a mayor número de servicios contratos por la empresa, mayor es el número de permanencia, del cual se reflejaría su vínculo de dependencia.

Cabe destacar la existencia de valores atípicos lo cual estaría vinculado a situaciones particulares en ciertos clientes, cual indicaría que, aunque existe tendencia general, también existen casos aislados.

En resumen, estas conclusiones enfatizan las diferencias de comportamiento entre las empresas grandes y pymes que aportan una perspectiva estratégica útil para orientar decisiones futuras, podemos decir que el análisis de los datos de las variables como el tiempo de permanencia, facturación y número de servicios contratados, nos ayuda a mejorar nuestras habilidades para identificar patrones y correlaciones, sino que también nos fortalece en el análisis de ellas.

5. Bibliografía

Andrew MacMillan, D. P. (n.d.). *libretexts*. Retrieved from libretexts:

[https://espanol.libretexts.org/Ingenieria/Ingenier%C3%ADa/Industrial_y_de_Sistemas/Libro%3A_Din%C3%A1mica_y_Control_de_Procesos_Qu%C3%ADmicos_\(Woolf\)/13%3A_Estad%C3%ADsticas_y_antecedentes_probabil%C3%ADsticos/13.01%3A_Estad%C3%ADsticas_b%C3%A1sicas](https://espanol.libretexts.org/Ingenieria/Ingenier%C3%ADa/Industrial_y_de_Sistemas/Libro%3A_Din%C3%A1mica_y_Control_de_Procesos_Qu%C3%ADmicos_(Woolf)/13%3A_Estad%C3%ADsticas_y_antecedentes_probabil%C3%ADsticos/13.01%3A_Estad%C3%ADsticas_b%C3%A1sicas)

https://www.it.uc3m.es/pablo/teoria-colas/labs/git-tr-lab01-es.html#1_introducci%C3%B3n_a_r. (2022, 02 10). *uc3m*. Retrieved from uc3m:
https://www.it.uc3m.es/pablo/teoria-colas/labs/git-tr-lab01-es.html#1_introducci%C3%B3n_a_r

jeanbragard. (n.d.). *github*. Retrieved from github:
<https://jeanbragard.github.io/PRACTICE/PracticasR2023bis.html#0-preparativos>

odiolaestadística. (n.d.). Retrieved from odiolaestadística:
<https://www.odiolaestadística.com/estadística-python/tipos-de-variables/>

Ortega, C. (2024). *questionpro*. Retrieved from questionpro :
<https://www.questionpro.com/blog/es/tabla-de-frecuencias/>

Ramírez-Alán, O. (2017, 10 09). *RPus*. Retrieved from RPus:
<https://rpubs.com/osoramirez/316691>

Rasilla, D. F. (02, 10 2023-). *personales unican*. Retrieved from personales unican:
https://personales.unican.es/rasillad/docencia/G14/TEMA_2/tablas_de_frecuencia.html

tutfg. (2024). Retrieved from tutfg: https://tutfg.es/ejemplos-conclusiones-tfg/?network=x&campaign=PMAX&group=&creative=&keyword=&device=c&matchtype=&gad_source=1&gclid=Cj0KCQjwvpy5BhDTARIsAHSilykytBMjJGxohjuxGNUBe1w7OuRwAqYcVAzpzUB9UwA3a0efzgBd7R8aAt_2EALw_wcB

6. Anexos

Anexo A. Para cargar datos y mostrarlos por variable

```
dir()
View("datos_grupo_12.txt")
datos<-read.table("datos_grupo_12.txt",header = TRUE) # Asignar variable en datos_grupo_12.txt
datos # Mostramos por consola
# Si no tenemos instalado openxlsx, instalamos el paquete, install.packages("openxlsx"), ejecutamos paquete library(openxlsx).
install.packages("openxlsx")
library("openxlsx")
write.xlsx(datos, file = "datos.xlsx", rowNames=TRUE,colNames=TRUE) # exportamos archivo de formato "txt" a una hoja de dato.
```

Anexo B. Explorar datos

```
dim(datos) # Para conocer el nº de observaciones y el nº de variables de la BBDD.
names(datos) # Para conocer los nombres de las variables, donde vemos que son "Tipo.de.empresa", "Número.de.Servicios" "Tiempo", "Facturado".
str(datos)
summary(datos) # Con esto veremos un resumen de la medida de cada una de las variables
attach(datos) # Desglosamos las variables.
```

Anexo C. Funciones

```
medidas.Disp<-function(nombre_tabla,decimales=2){
tf.discretos<-function(nombre_tabla,decimales=2){
tf.continuos<-function(nombre_tabla,observ,decimales=2){

medidas.Disp<-function(nombre_tabla,decimales=2){
  resultado<-try({
    media<-round(mean(nombre_tabla),digits = decimales)
    desviacion<-round(sd(nombre_tabla),digits = decimales)
    varianza<-round(var(nombre_tabla),digits = decimales)
    coeficiente.vr<-round((desviacion/media)*100,digits = decimales)

    tabla<-data.frame(max(nombre_tabla)-min(nombre_tabla),median(nombre_tabla),media,varianza,desviacion,coeficiente.vr)
    names(tabla)<-c("Rango","Mediana","Media","Varianza","Desviación estadar","Coeficiente de variación")
    return(tabla)
  },silent = FALSE)
  if(inherits(resultado,"try-error")){
    cat("Ocurre un error verificar: \n")
  }else{
    cat(resultado)
  }
}

tf.discretos<-function(nombre_tabla,decimales=2){
  resultado<-try({
    if(inherits(resultado,"try-error")){
      cat("Ocurre un error verificar \n")
    }else{
    }
  },silent = FALSE)
  if(inherits(resultado,"try-error")){
    cat("Ocurre un error verificar \n")
  }else{
  }
}

tf.continuos<-function(nombre_tabla,observ,decimales=2){
  resultado<-try(
  {
    n.clases<-ceiling(1+3.222*log10(observ))
    anch.inter<-round((max(nombre_tabla)-min(nombre_tabla))/n.clases,digits = decimales)
    limites<-min(nombre_tabla)+c(0:n.clases)*anch.inter
    cortes<-cut(nombre_tabla,breaks =limites,include.lowest = TRUE,dig.lab = 3+decimales)
    tabla<-table(cortes)

    marca.clases<-numeric(n.clases)
    for(i in 1:n.clases){
      resultados<-round((2*limites[i]+anch.inter)/2,digits = 2)
      marca.clases[i]<-resultados
    }
    tf<-as.data.frame(tabla)
    tfComp1<-transform(tf,
      Xi=marca.clases,
      FreqAc=cumsum(tf$Freq),
      Relat=round(prop.table(tf$Freq),digits = decimales),
      RelatAc=round(cumsum(prop.table(tf$Freq)),digits = decimales)
    )
    #rm(anch.inter,limites,cortes,n.clases,marca.clases,i)
    return(tfComp1)
  },silent=FALSE)
  if(inherits(resultado,"try-error")){
    cat("Ocurrió un error, vuelve a verificar: \n")
  }
  cat(resultado)
}
}
```


Anexo D. Análisis Descriptivo Univariante

```
# 4 Análisis Descriptivo Univariante
# Variable cualitativa:
#-----
# 1. Tipo de empresa
tb.TipoEmp<-tf.discretos(Tipo.de.empresa) # Frecuencia relativa, donde el 36...% aproximadamente son empresas "grandes" y el resto (63..%) son empresas "pyme".
tb.TipoEmp
write.xlsx(tb.TipoEmp,file = "tb.TipoEmp.xlsx")

# Gráfico
par(mfrow=c(1,2)) # Configuramos parámetros del gráfico
barplot(tabla.tb.tipo.empresa,main="Tipo de empresa", col = "lightblue",ylab = "Frecuencias absolutas", xlab = "Empresa")
pie(tabla.tb.tipo.empresa,main="Tipo de empresa",ylab = "Frecuencias absolutas", xlab = "Empresa") # Diagrama de círculo

# Variables cuantitativas:
#-----
# 1. Numero de servicios
# Tabla de Frecuencias:
tb.NumSer<-tf.discretos(Número.de.Servicios)
tb.NumSer
write.xlsx(tb.NumSer,file = "tb.NumSer.xlsx")

# Medidas de dispersión:
md.NumSer<-medidas.Disp(Número.de.Servicios)
cat("Medidas de dispersión de Número de Servicios: \n")
md.NumSer

# Distribución, veremos como se están distribuyendo los datos
install.packages("moments") # Instalar el paquete moments si no lo tienes
library(moments) # Cargar el paquete
skewness(Número.de.Servicios) # 0.01985621 es >= 0, lo que significa que es prácticamente simétrica, lo veremos en un gráfico para aclarar más.
plot(density(Número.de.Servicios), main = "Densidad del Número de Servicios") # Curva de densidad

# Gráficos
hist(Número.de.Servicios, breaks = "Sturges", freq = TRUE,main="Histograma de Nº de servicios", xlab = "Nº de servicios", ylab = "Frecuencias Absolutas")
hist(Número.de.Servicios, breaks = "Sturges", freq = FALSE,main="Histograma de Nº de servicios", xlab = "Nº de servicios", ylab = "Frecuencias Relativas")
boxplot(Número.de.Servicios,main="Diagrama de Nº de servicios",ylab="Valores",col = "lightblue") # Resaltaremos los cuartiles de la variable, así como la presencia de valores atípicos (outliers).

# 2. Tiempo
# Tabla de Frecuencias
tb.Tiempo<-tf.contínuos(Tiempo,207)
tb.Tiempo
write.xlsx(tb.Tiempo,file = "tb.Tiempo.xlsx")

# Medidas de dispersión:
md.Tiempo<-medidas.Disp(Tiempo)
cat("Medidas de dispersión del tiempo: \n")
md.Tiempo

# Distribución
skewness(Tiempo) # 0.01360188, se aproxima a una distribución simétrica.
plot(density(Tiempo),main="Tiempo")

# Gráficos
hist(Tiempo, breaks = "Sturges", freq = TRUE,main="Histograma de tiempo", xlab = "Tiempo", ylab = "Frecuencias Absolutas")
hist(Tiempo, breaks = "Sturges", freq = FALSE,main="Histograma de tiempo", xlab = "Tiempo", ylab = "Frecuencias Relativas")
boxplot(Tiempo,main="Diagrama de tiempo",ylab="Valores",col = "lightblue")

#-----
# 3. Facturado
# Tabla de Frecuencias
tb.Factur<-tf.contínuos(Facturado,207)
tb.Factur
write.xlsx(tb.Factur,file = "tb.Factur.xlsx") # Importar a excel

# Medidas de dispersión:
md.Facturado<-medidas.Disp(Facturado)
cat("Medidas de dispersión de Facturado: \n")
md.Facturado

# Distribución
skewness(Facturado) # 0.1519434 (Simétrica).
plot(density(Facturado))

# Gráficos
hist(Facturado, breaks = "Sturges", freq = TRUE,main="Histograma de Facturado", xlab = "Facturado", ylab = "Frecuencias Absolutas")
boxplot(Facturado,main="Diagrama de Facturado",ylab="Valores",col = "lightblue")
par(mfrow=c(1,1)) # Volvemos a la configuración
```

Anexo E. Análisis de la Variables Tiempo

```
boxplot(Tiempo ~ Tipo.de.empresa, data=datos, main="Tiempo por Tipo de Empresa", ylab="Tiempo (meses)", xlab = "Tipo de Empresa", col = c("lightblue", "lightgreen")) # Boxplot del Tiempo según el Tipo de empresa
boxplot(Tiempo ~ Número.de.Servicios, data=datos, main="Tiempo por Número de Servicios", ylab="Tiempo (meses)", xlab="Número de Servicios", col="lightblue") # Boxplot del Tiempo por Número de Servicios

media.tiempo.empresa <- aggregate(Tiempo ~ Tipo.de.empresa, data=datos, FUN=mean)
barplot(media.tiempo.empresa$Tiempo, # Gráfico de barras: Media del Tiempo según Tipo de Empresa
names.arg = media.tiempo.empresa$Tipo.de.empresa,
col = c("lightblue", "lightgreen"),
main = "Media del Tiempo por Tipo de Empresa",
xlab = "Tipo de Empresa",
ylab = "Media del Tiempo (meses)")

media.tiempo.servicios <- aggregate(Tiempo ~ Número.de.Servicios, data=datos, FUN=mean) # Gráfico de barras para la Media del Tiempo por Número de Servicios
barplot(media.tiempo.servicios$Tiempo,
names.arg = media.tiempo.servicios$Número.de.Servicios,
col = "lightblue",
main = "Tiempo Promedio por Servicios Contratados",
xlab = "Número de Servicios",
ylab = "Tiempo Promedio (meses)")

# Calcular la correlación entre el tiempo, tipo de servicio, numero de servicio:
tipoEmp.num<-as.numeric(Factor(Tipo.de.empresa))
cor.Tien.Tiempo <- cor(Tiempo,tipoEmp.num,use = "complete.obs",method = "spearman") # x, y deben ser numericas!!
cor.Tien.numSer <- cor(Tiempo,Número.de.Servicios,use = "complete.obs",method = "spearman")

# Diagrama de dispersión para visualizar la relación entre el tiempo, tipo de servicio, numero de servicio.
plot(tipoEmp.num, Tiempo, main = "Diagrama de Dispersión: Tiempo vs Tipo de Empresa", xlab="Tipo de empresa",col = "lightblue", pch = 19)
abline(lm(Tiempo ~ tipoEmp.num, data=datos), col="red") # Añadir línea de tendencia al gráfico de dispersión

plot(Número.de.Servicios, Tiempo, main = "Diagrama de Dispersión: Tiempo vs Número de Servicios",xlab="Número de servicios", col = "lightblue", pch = 19)
abline(lm(Tiempo ~ Número.de.Servicios, data=datos), col="red") # Añadir línea de tendencia al gráfico de dispersión
```

Anexo F. Análisis de la Variable Facturado

```
boxplot(Facturado ~ Tipo.de.empresa, data = datos, main="Facturado por tipo de empresa", ylab = "Facturado", xlab = "Tipo de empresa", col="lightblue")
boxplot(Facturado ~ Número.de.Servicios, data = datos, main="Facturado por Número de Servicios", ylab = "Facturado", xlab = "Número de servicio", col="lightblue")

media.fact.tipoEmp<-aggregate(Facturado ~ Tipo.de.empresa,FUN = mean)
barplot(media.fact.tipoEmp$Facturado,names.arg = media.fact.tipoEmp$Tipo.de.empresa,main="Facturado Promedio por Tipo de servicio", ylab = "Facturado",xlab ="Número de servicios", col="lightblue")

media.fact.numServ<-aggregate(Facturado ~ Número.de.Servicios,FUN = mean)
barplot(media.fact.numServ$Facturado,names.arg = media.fact.numServ$Número.de.Servicios, main = "Facturado Promedio por Número de servicios", ylab = "Facturado",xlab ="Número de servicios", col="lightblue")
```

Anexo G. Correlación entre el facturado, tipo de servicio, número de servicio

```
cor.Fact.tiEmpr <- cor(x = Facturado,y=tipoEmpr.num,use = "complete.obs",method = "spearman") # -0.4100544
cor.Fact.NumServ <- cor(x = Facturado,y=Número.de.Servicios,use = "complete.obs",method = "spearman") # 0.7041209

plot(tipoEmpr.num,Facturado,main = "Diagrama de dispersión: Facturado vs Tipo de Empresa", xlab="Tipo de empresa",col = "lightblue", pch = 19)
abline(lm(Facturado ~ tipoEmpr.num, data=datos),col="red")

plot(Número.de.Servicios,Facturado,main = "Diagrama de dispersión: Facturado vs Número de servicios", xlab="Número de servicios",col = "lightblue", pch = 19)
abline(lm(Facturado ~ Número.de.Servicios, data=datos),col="red")
```

Anexo H. Relación Bivariante entre Tiempo y Facturado

```
plot(Tiempo, Facturado,main = "Diagrama de dispersión entre tiempo y facturado", col="blue")
abline(lm(Facturado-Tiempo, data = datos), col="red")
cor(Facturado,Tiempo,use="complete.obs",method = "spearman") # 0.8987834
reg.Fact.Tiem<-lm(Facturado-Tiempo,data = datos)
```

```
> summary(reg.Fact.Tiem)

Call:
lm(formula = Facturado ~ Tiempo, data = datos)

Residuals:
    Min       1Q   Median       3Q      Max
-5281  -1798   -141   1723   6925

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -645.651    563.074  -1.147   0.253
Tiempo       92.103      3.115   29.564 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2387 on 205 degrees of freedom
Multiple R-squared:  0.81,    Adjusted R-squared:  0.8091
F-statistic: 874 on 1 and 205 DF, p-value: < 2.2e-16
```

Anexo I. Análisis con Distinción de Tipo de empresa

```
# Instalación paquete dplyr
install.packages("dplyr")
library(dplyr)

# Tabla de medidas variable Tiempo
medidas_tiempo <- datos %>%
  group_by(Tipo.de.empresa) %>%
  summarise(
    Media_Tiempo = mean(Tiempo),
    Mediana_Tiempo = median(Tiempo),
    Varianza_Tiempo = var(Tiempo),
    Desviacion_Tiempo = sd(Tiempo)
  )
medidas_tiempo

# Tabla de medidas variable Facturado
medidas_facturado <- datos %>%
  group_by(Tipo.de.empresa) %>%
  summarise(
    Media_Facturado = mean(Facturado),
    Mediana_Facturado = median(Facturado),
    Varianza_Facturado = var(Facturado),
    Desviacion_Facturado = sd(Facturado),
  )
medidas_facturado
```