



Universidad  
Andrés Bello

Actividad

Evaluativa #1

Grupo 5

Martín

Fernández

Tomás Moya

Wesly

Ocampo

Alan Tovar



Universidad  
Andrés Bello

# Actividad Evaluativa #1

## Grupo 5

### Facultad de Ingeniería

Martín Fernández

Tomás Moya

Wesly Ocampo

Alan Tovar

Universidad Andrés Bello  
Ingeniería Civil Informática  
Ingeniería Civil Industrial

Septiembre 2022



Universidad  
Andrés Bello

Actividad  
Evaluativa #1  
Grupo 5

Martín  
Fernández  
Tomás Moya  
Wesly  
Ocampo  
Alan Tovar

# Introducción

- La Importancia del Preprocesamiento.
- Administración de Datos.
- Análisis Exploratorio de Datos.



Figure 1: ...



Universidad  
Andrés Bello

Actividad

Evaluativa #1

Grupo 5

Martín  
Fernández  
Tomás Moya  
Wesly  
Ocampo  
Alan Tovar

## Renta de Bicicletas (rentBike)

- 5856 observaciones.
- 14 variables:
  - Date -Rented\_Bike\_Count
  - Hour
  - Temperature
  - Humidity
  - Windspeed
  - Visibility
  - Dew\_point\_temperature
  - Solar\_Radiation
  - Rainfall
  - Snowfall
  - Seasons
  - Holiday
  - Functioning\_Day



Universidad  
Andrés Bello

Actividad  
Evaluativa #1  
Grupo 5

Martín  
Fernández  
Tomás Moya  
Wesly  
Ocampo  
Alan Tovar

# Tipos de Variables

## Variables: Cualitativas, Cuantitativas (Enteros y Numéricas)

variables	types	missing_count	missing_percent	unique_count	unique_rate
Date	character	0	0	244	0.0416666667
Rented_Bike_Count	integer	0	0	2,125	0.3628756831
Hour	integer	0	0	24	0.0040983607
Temperature	numeric	0	0	398	0.0679644809
Humidity	integer	0	0	90	0.0153688525
Windspeed	numeric	0	0	62	0.0105874317
Visibility	integer	0	0	1,597	0.2727117486
Dew_point_temperature	numeric	0	0	421	0.0718920765
Solar_Radiation	numeric	0	0	344	0.0587431694
Rainfall	numeric	0	0	56	0.0095628415
Snowfall	numeric	0	0	27	0.0046106557
Seasons	character	0	0	3	0.0005122951
Holiday	character	0	0	2	0.0003415301
Functioning_Day	character	0	0	2	0.0003415301



Universidad  
Andrés Bello

Actividad

Evaluativa #1

Grupo 5

Martín

Fernández

Tomás Moya

Wesly

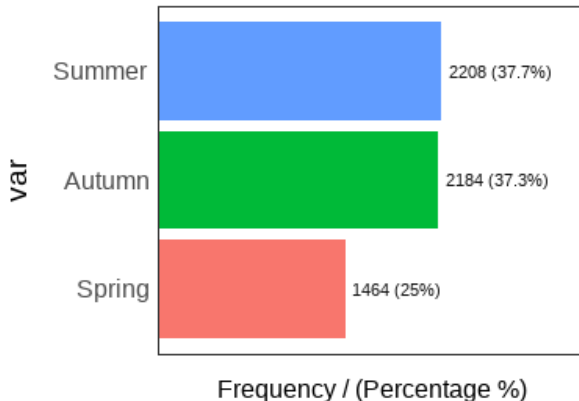
Ocampo

Alan Tovar

# Gráfico de Barras de la variable

Representación gráfica de la frecuencia en Seasons.

Obs: No existe la Categoría "Winter"





Universidad  
Andrés Bello

Actividad

Evaluativa #1

Grupo 5

Martín

Fernández

Tomás Moya

Wesly

Ocampo

Alan Tovar

# Resumen de medidas estadísticas de todas las variables, en una tabla.

- Mínimos, Máximos, Medias, Cuartiles, Datos Atípicos. . .

variables	min	Q1	mean	median	Q3	max	zero	minus	outlier
Rented_Bike_Count	0	327.75	905.8317964	813.00	1,298.50	3,556.00	295	0	55
Hour	0	5.75	11.5000000	11.50	17.25	23.00	244	0	0
Temperature	-3	12.90	19.1865949	19.90	25.20	39.40	1	21	0
Humidity	0	46.00	61.2187500	61.00	77.00	98.00	17	0	0
Windspeed	0	0.90	1.6269467	1.50	2.20	7.40	48	0	100
Visibility	27	1,003.00	1,470.7764686	1,731.00	2,000.00	2,000.00	0	0	0
Dew_point_temperature	-19	4.10	10.7083333	11.40	18.70	27.20	43	726	3
Solar_Radiation	0	0.00	0.6651178	0.05	1.17	3.52	2,664	0	196
Rainfall	0	0.00	0.2015198	0.00	0.00	35.00	5,392	0	464
Snowfall	0	0.00	0.0210041	0.00	0.00	8.80	5,805	0	51



Universidad  
Andrés Bello

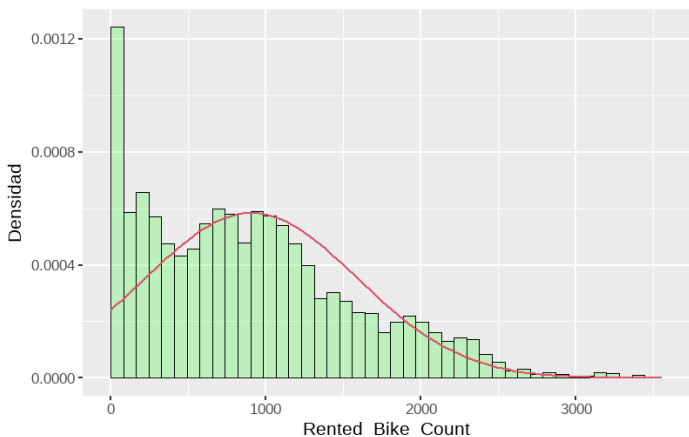
Actividad

Evaluativa #1

Grupo 5

Martín  
Fernández  
Tomás Moya  
Wesly  
Ocampo  
Alan Tovar

# Análisis de variable numérica con histograma de frecuencias y curva de densidad.



**Figure 2:** Barplot Variable “Rented\_Bike\_Count”



Universidad  
Andrés Bello

Actividad  
Evaluativa #1  
Grupo 5

Martín  
Fernández  
Tomás Moya  
Wesly  
Ocampo  
Alan Tovar

# Análisis de normalidad de variable numérica con gráfico Q-Q

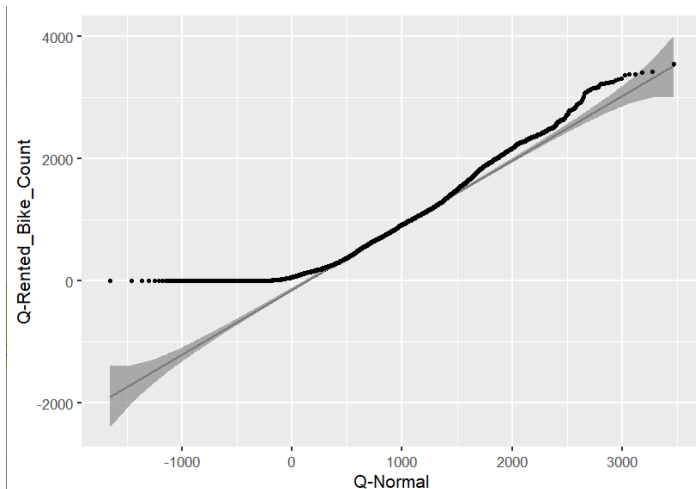


Figure 3: Q-Q



## Prueba de Normalidad

[illegible]

Test de kolmogorov-smirnov:

```
data: datos$Rented Bike Count
```

$$D = 0.092264$$

p-value < 2.2e-16

alternative hypothesis:

two-sided



Universidad  
Andrés Bello

Actividad

Evaluativa #1

Grupo 5

Martín

Fernández

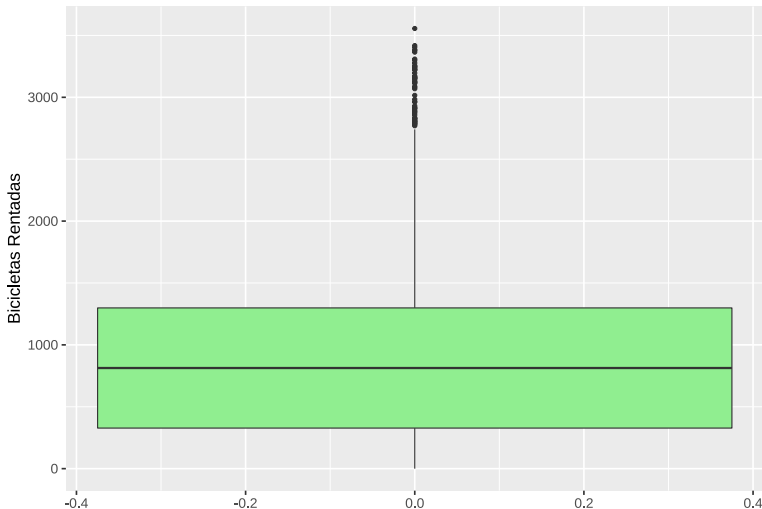
Tomás Moya

Wesly

Ocampo

Alan Tovar

# Identificación de datos atípicos de la variable numérica con uso de boxplot.





Universidad  
Andrés Bello

Actividad

Evaluativa #1

Grupo 5

Martín

Fernández

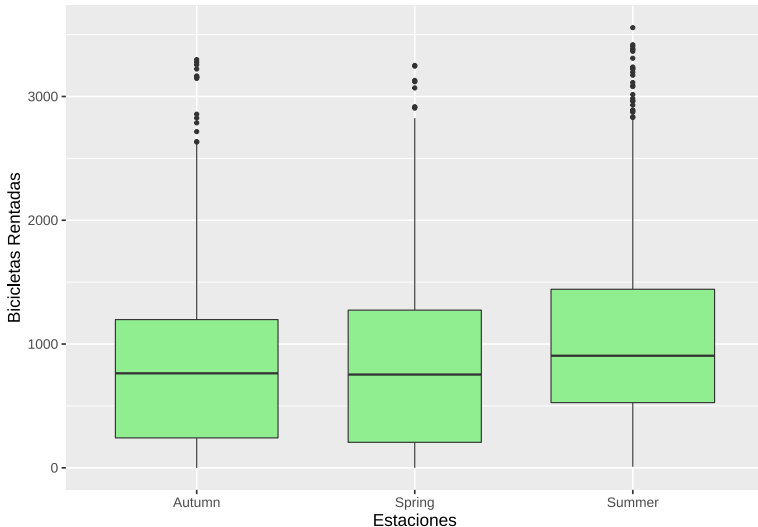
Tomás Moya

Wesly

Ocampo

Alan Tovar

# Uso de boxplot para relacionar la variable numérica con la variable entera o categórica del análisis anterior.





Universidad  
Andrés Bello

Actividad

Evaluativa #1

Grupo 5

Martín

Fernández

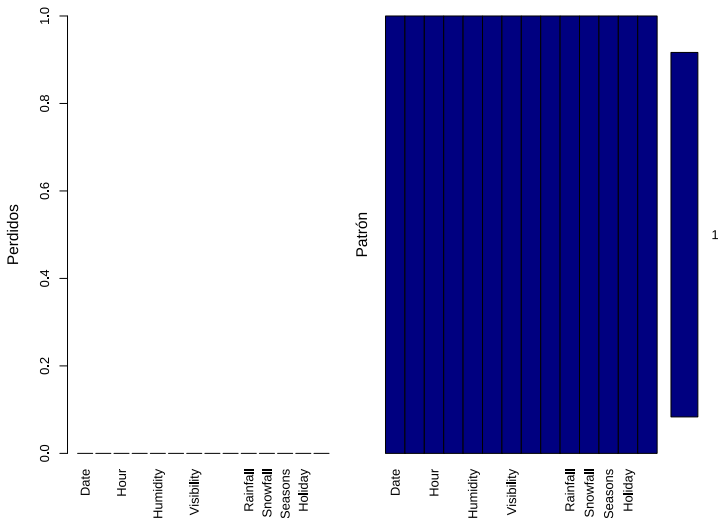
Tomás Moya

Wesly

Ocampo

Alan Tovar

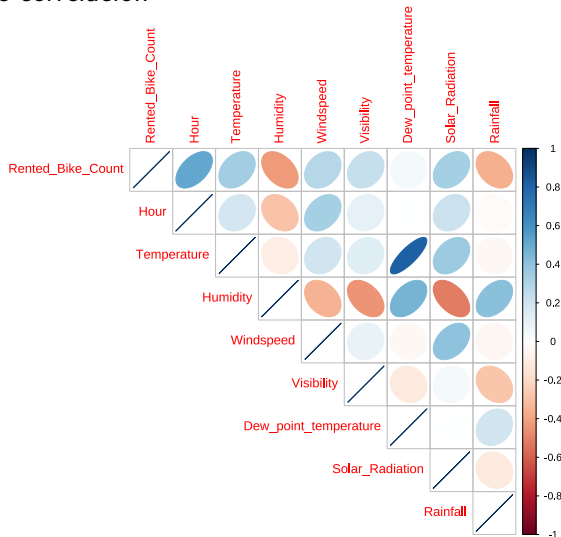
# Gráfico que visualice la proporción de datos faltantes.



##

## Representación gráfica de la matriz de correlación.

### Matriz de correlación





Universidad  
Andrés Bello

Actividad

Evaluativa #1

Grupo 5

Martín

Fernández

Tomás Moya

Wesly

Ocampo

Alan Tovar

# Conclusiones

- Alto nivel de correlación entre la variables.
- Ningún dato faltante.
- La variable Rented bike count no está normalizada y es asimétrica positiva.
- Datos atípicos.
- Resultados del análisis en la venta de Bicicletas.