



Universidad  
Andrés Bello®

Actividad Evaluativa #1

Grupo 5

Universidad Andrés Bello

Facultad de Ingeniería

Ingeniería Civil Informática

Ingeniería Civil Industrial

Martín Fernández 1<sup>1</sup>, Tomás Moya 2, Wesly Ocampo 3<sup>1</sup>, Alan Tovar 4<sup>1</sup>

---

### Abstract

Describir brevemente en qué consiste este primer análisis de los datos, e incluir el objetivo del estudio. (máximo 150 palabras)

---

---

\*Corresponding author

Email addresses: `nombre@uandresbello.edu` (Martín Fernández 1),  
`nombre@uandresbello.edu` (Tomás Moya 2), `nombre@uandresbello.edu` (Wesly Ocampo 3),  
`nombre@uandresbello.edu` (Alan Tovar 4)

Tabla de contenidos

Índice de figuras y tablas

## Introducción

El preprocesamiento de datos y la administración de los mismos nos permiten la recolección de datos de distintas fuentes, el tratamiento de filas y cabeceras (headers o columns). Con esto podemos hacer una limpieza adecuada, eliminar errores, corregir inconsistencias y aumentar la calidad de la minería de datos, una correcta gestión de datos también nos sirve para acceder a diferentes datos de una forma más fácil, así obtenemos información estadística y comparativa que permite una correcta toma de decisiones y abarcar de mejor manera los distintos problemas empresariales.

Utilizar referencias para explicar la importancia del análisis exploratorio de datos. (Ejemplo para citar las referencias [1])

Descripción del conjunto de datos (número de observaciones, número de variables, cuáles son las variables involucradas en el estudio, tipos de variables)

Editar este código, debe incluir las librerías que necesita, y cargar los datos:

Table 1: Resumen de medidas estadísticas.

variables	min	Q1	mean	median	Q3	max	zero	minus	outliers
Rented_Bike_Count	0	327.75	905.83	813.00	1298.50	3556.00	295	0	53
Hour	0	5.75	11.50	11.50	17.25	23.00	244	0	0
Temperature	-3	12.90	19.19	19.90	25.20	39.40	1	21	0
Humidity	0	46.00	61.22	61.00	77.00	98.00	17	0	0
Windspeed	0	0.90	1.63	1.50	2.20	7.40	48	0	100
Visibility	27	1003.00	1470.78	1731.00	2000.00	2000.00	0	0	0
Dew_point_temperature	-19	4.10	10.71	11.40	18.70	27.20	43	726	3
Solar_Radiation	0	0.00	0.67	0.05	1.17	3.52	2664	0	190
Rainfall	0	0.00	0.20	0.00	0.00	35.00	5392	0	464
Snowfall	0	0.00	0.02	0.00	0.00	8.80	5805	0	53

## Desarrollo

### *Resumen de medidas estadísticas*

Resumen de medidas estadísticas de rentBike (en una tabla), e interpretación.

*Análisis de la variable Seasons*

Análisis de una variable tipo entera o categórica, representación gráfica de la frecuencia de cada categoría

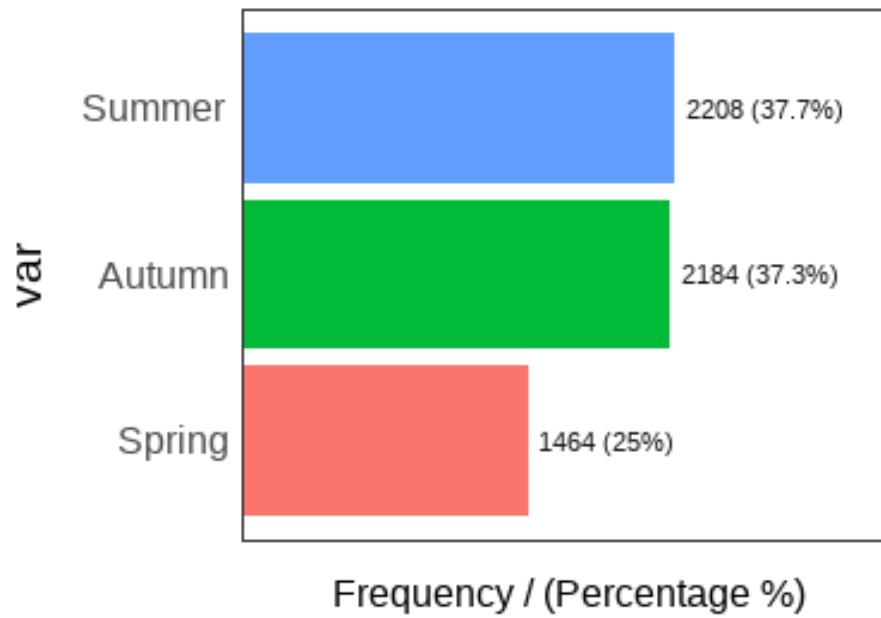


Figure 1: Histograma de Seasons.

Table 2: Tabla de frecuencias de Bicicletas Rentadas Rented Bike Count.

variables	levels	N	freq	ratio	rank
Rented_Bike_Count	(140,160]	641	91	14.20	1
Rented_Bike_Count	(20,40]	641	87	13.57	2
Rented_Bike_Count	(160,180]	641	82	12.79	3
Rented_Bike_Count	(0,20]	641	80	12.48	4
Rented_Bike_Count	(120,140]	641	75	11.70	5
Rented_Bike_Count	(40,60]	641	72	11.23	6
Rented_Bike_Count	(60,80]	641	56	8.74	7
Rented_Bike_Count	(100,120]	641	54	8.42	8
Rented_Bike_Count	(80,100]	641	44	6.86	9

#### *Análisis de la variable Bicicletas Rentadas Rented Bike Count*

Análisis de una variable tipo numérica (tabla de frecuencias, histograma de frecuencias y curva de densidad) y análisis de normalidad (Q-Q plot y prueba de normalidad)

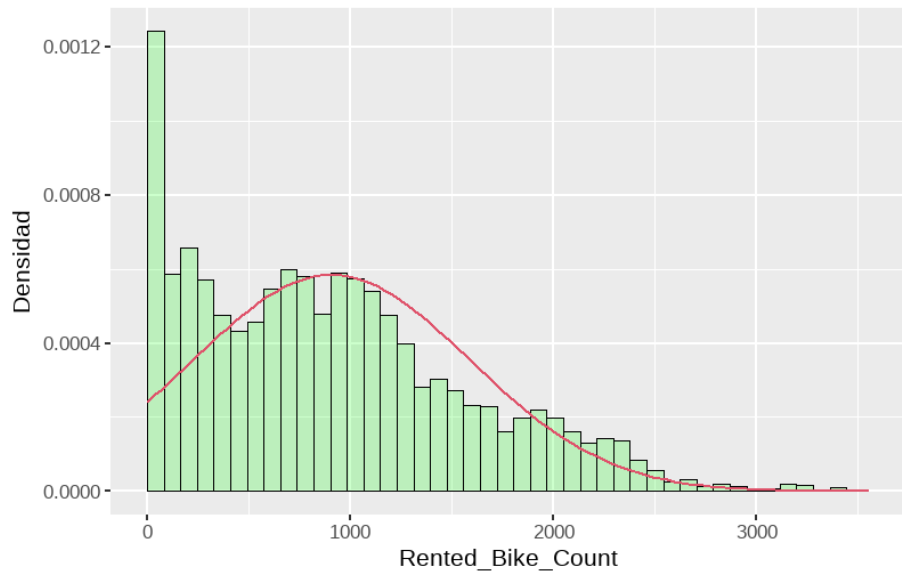


Figure 2: Histograma con curva de densidad de Bicicletas Rentadas Rented Bike Count.

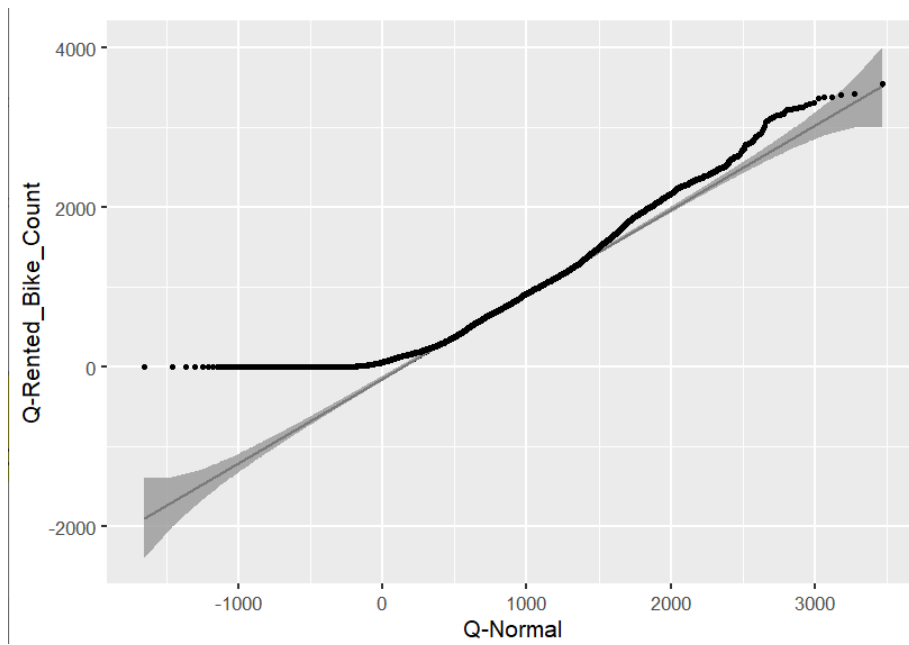


Figure 3: Q-Q plot de Bicicletas Rentadas Rented Bike Count.

### Valores Atípicos

Identificación de valores atípicos (boxplot variable numérica (2.3), boxplot relación variable numérica (2.3) con variable entera o categórica (2.2))

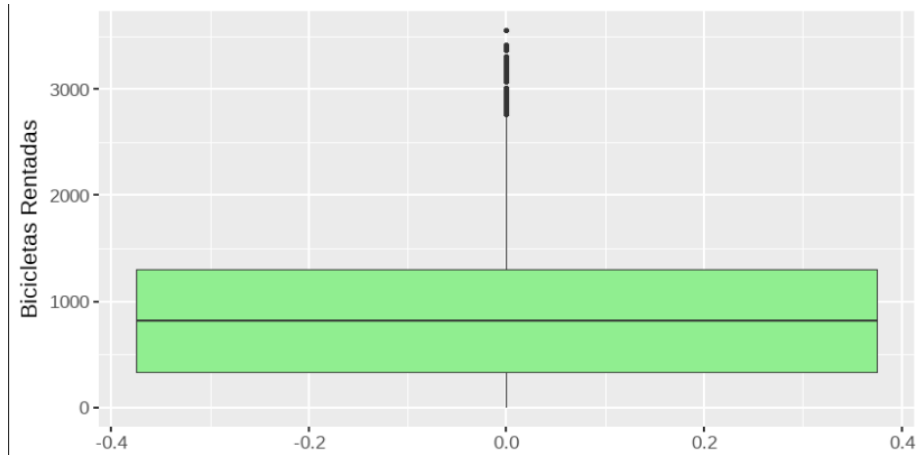


Figure 4: Datos Atípicos de Bicicletas Rentadas Rented Bike Count.

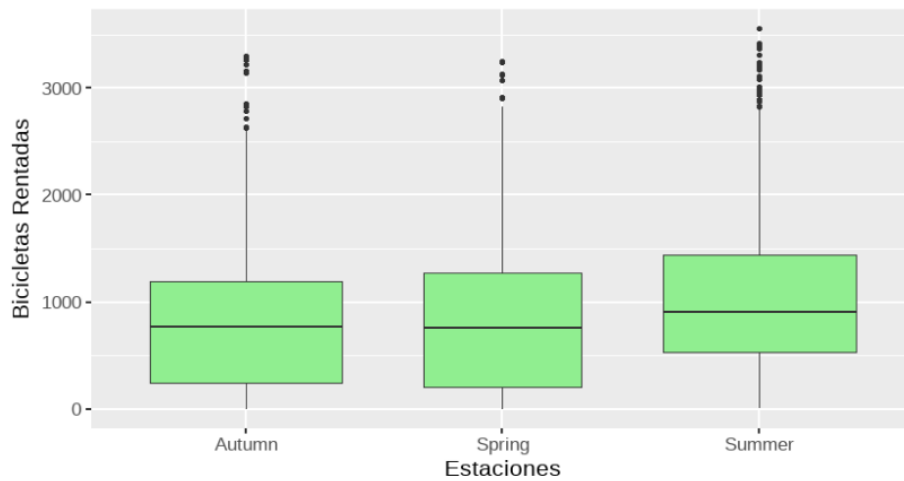


Figure 5: Boxplot de la variable Seasons.



### *Datos Faltantes*

Determinar proporción de datos faltantes

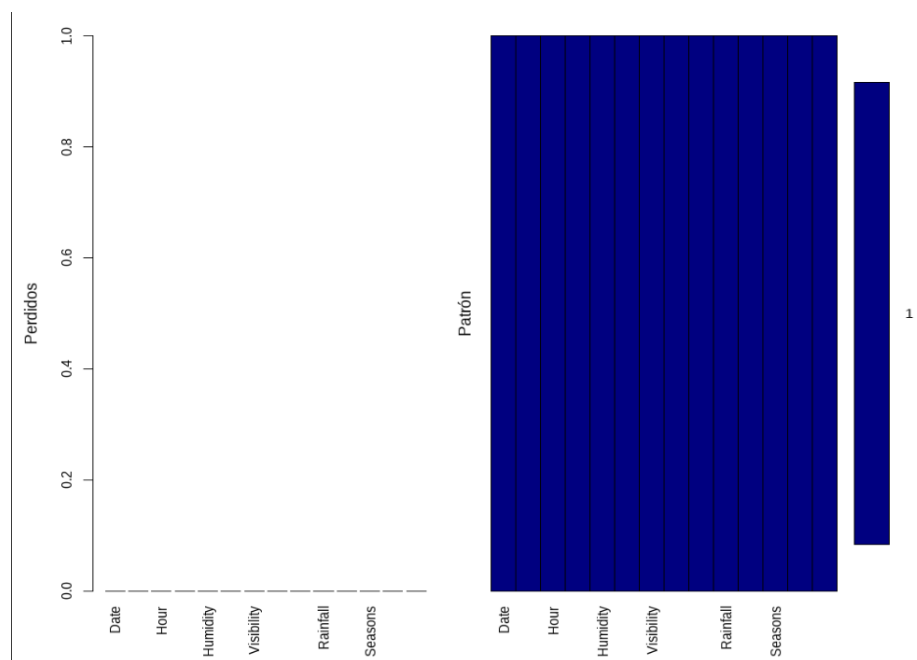


Figure 6: Gráfico de Datos faltantes.

### *Análisis de Correlación*

Análisis de correlación, representación de matriz de correlación.

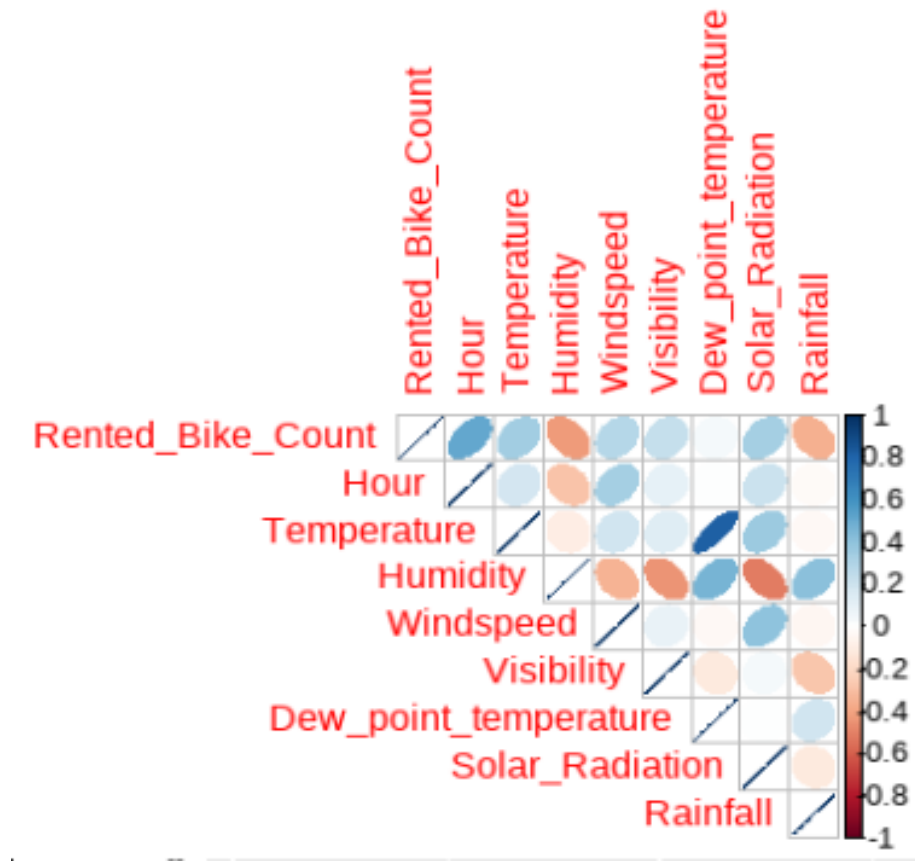


Figure 7: Matriz de Correlación.

### **Conclusiones**

Resuma las principales conclusiones de cada análisis realizado como parte del desarrollo.

### **Referencias Bibliográficas**

Referencias Bibliográficas {#references .unnumbered}

## Anexos

*fig*

```
Ejemplo para hacer la tabla (Ver Tabla 1)
setwd("D:\\UNAB\\MINERIA DE DATOS\\AE1") datos <- read.csv("rentBike_data.csv",sep
= ",",header = TRUE) diagnose_numeric(datos) %>% flextable()
funModeling::freq(datos$Seasons)
=====
```

## References

- [1] Foster Provost and Tom Fawcett. Data science and its relationship to big data and data-driven decision making. *Big data*, 1(1):51–59, 2013.