

PROYECTO 1

BICICLETAS COMPARTIDAS

DATA TRIO
CATALINA GÓMEZ
VIVIANA LARA
JUAN DAVID RIOS

RESUMEN DE CONTENIDOS

CONTEXTO

TABLERO

ANÁLISIS DE NEGOCIO

DESPLIEGUE

LIMPIEZA DATOS

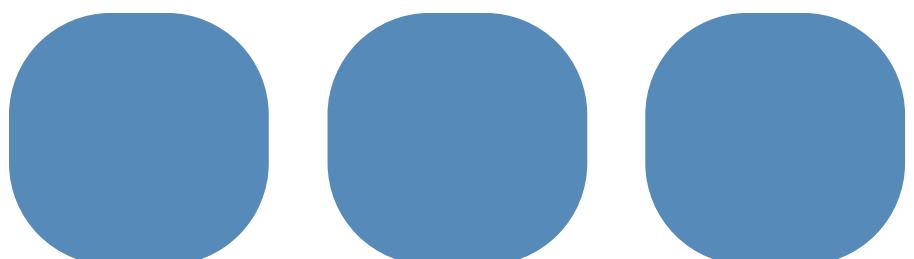
CONCLUSIONES

EXPLORACIÓN DATOS

BICICLETAS COMPARTIDAS SEÚL



- El operador del sistema "Ddareungi" debe asegurar un buen nivel de servicio durante todo el año.
- Necesidad de contratar personal, disponer de vehículos de transporte y equipos de mantenimiento.
- Clima continental húmedo: inviernos fríos (-5°C) y secos, veranos calurosos (hasta 25°C) y húmedos.





- Factores climáticos que afectan el uso: radiación solar intensa en verano, visibilidad reducida por niebla y contaminación.
- Infraestructura ciclista limitada en algunas áreas, lo que demanda mejoras para la seguridad vial.
- Adaptaciones necesarias para mantener y fomentar el uso de bicicletas durante todo el año.

CLIENTE

ÁREA DE OPERACIONES



ANÁLISIS DE NEGOCIO

¿Cómo varía la demanda de bicicletas a lo largo del día?

¿Cómo afectan las condiciones meteorológicas el uso del sistema de bicicletas compartidas?

¿Cuál de las condiciones meteorológicas tiene un mayor impacto en la demanda?

¿Cuál sería un valor esperado para la demanda de bicicletas para enero del 2019?

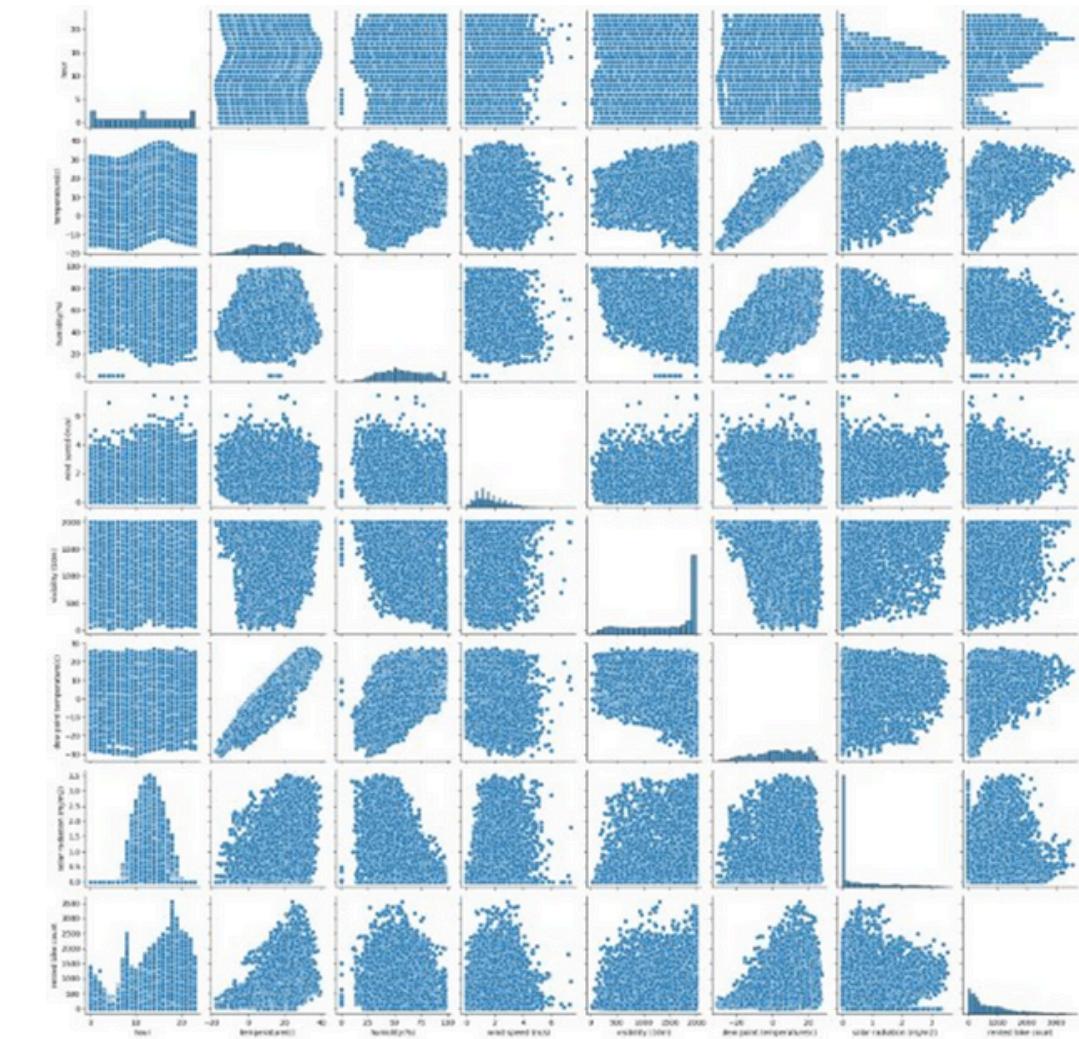


EXPLORACIÓN Y LIMPIEZA DE DATOS

Analisis de datos

Estadisticas descriptivas

	date	rented bike count	hour	temperature(c)	humidity(%)	wind speed (m/s)	visibility (10m)	dew point temperature(c)	solar radiation (mj/m2)	rainfall(mm)	snowfall (cm)
count	8760	8760.000000	8760.000000	8760.000000	8760.000000	8760.000000	8760.000000	8760.000000	8760.000000	8760.000000	8760.000000
mean	2018-05-31 23:59:59.999999	704.602055	11.500000	12.882922	58.226256	1.724909	1436.825799	4.073813	0.569111	0.148687	0.075068
min	2017-12-01 00:00:00	0.000000	0.000000	-17.800000	0.000000	0.000000	27.000000	-30.600000	0.000000	0.000000	0.000000
25%	2018-03-02 00:00:00	191.000000	5.750000	3.500000	42.000000	0.900000	940.000000	-4.700000	0.000000	0.000000	0.000000
50%	2018-06-01 00:00:00	504.500000	11.500000	13.700000	57.000000	1.500000	1698.000000	5.100000	0.010000	0.000000	0.000000
75%	2018-08-31 00:00:00	1065.250000	17.250000	22.500000	74.000000	2.300000	2000.000000	14.800000	0.930000	0.000000	0.000000
max	2018-11-30 00:00:00	3556.000000	23.000000	39.400000	98.000000	7.400000	2000.000000	27.200000	3.520000	35.000000	8.800000
std	Nan	644.997468	6.922582	11.944825	20.362413	1.036300	608.298712	13.060369	0.868746	1.128193	0.436746

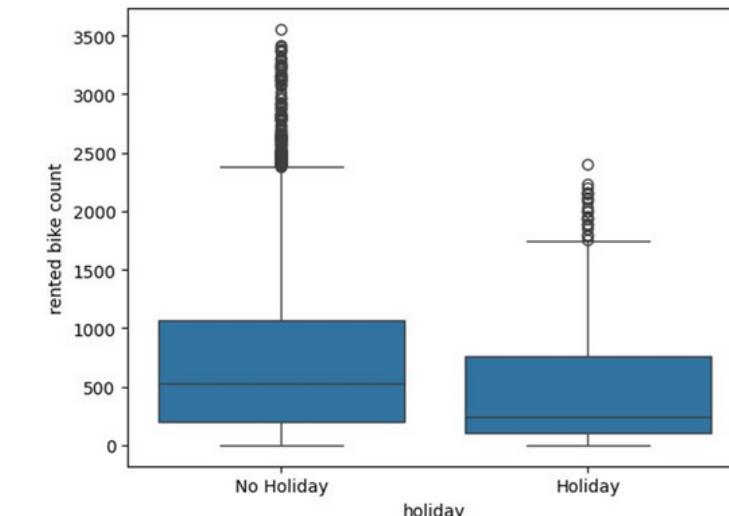


Se descartaron 5 variables de respuesta por diversas razones:

- rainfall(mm) : presenta el ~94% de los registros en 0
- snowfall (cm) : presenta ~95% de los registros en 0
- dew point temperature: presenta una correlacion del 91% con temperature
- holiday: solo 17 dias de observación corresponden a holiday y no se observa un comportamiento diferencial sobre el target
- date: la variable seasons logra explicar de manera mas resumida el efecto de la fecha.

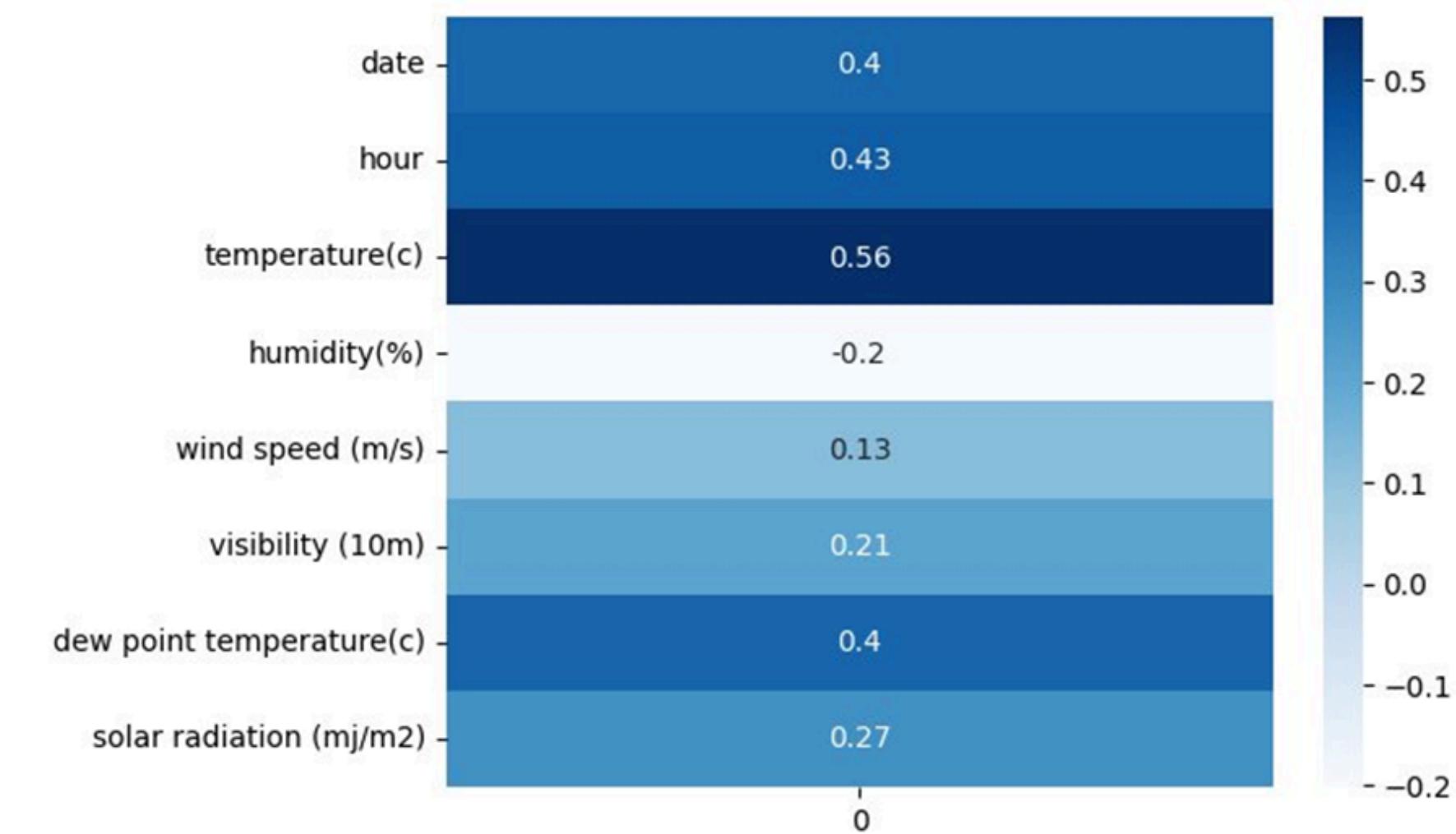
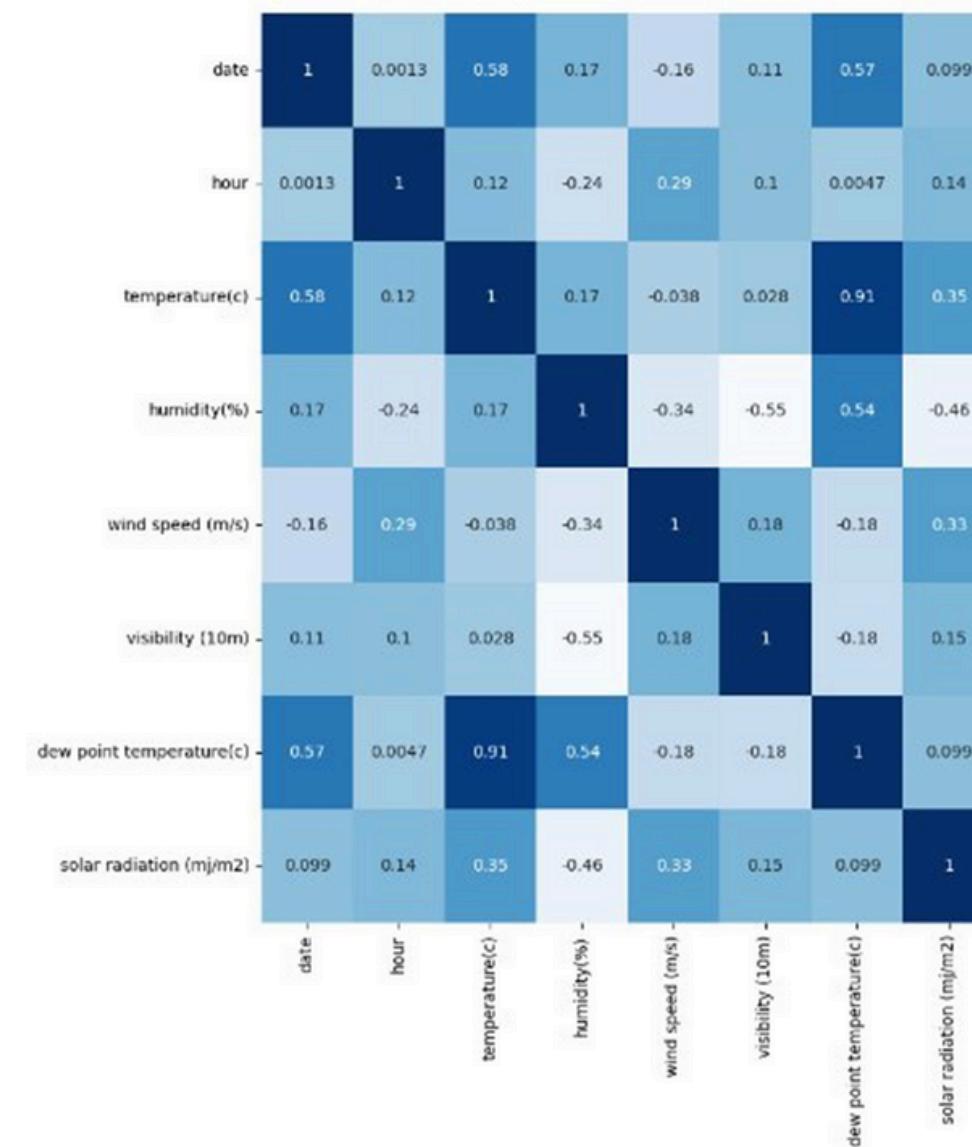
Se observa gran interacción entre las variables explicativas y de respuesta, especialmente se observa que la mayoría de variables relacionadas con las características ambientales tienen un efecto sobre la variable de respuesta.

Igualmente, se observa un gran comportamiento estacional en la serie de tiempo, lo cual nos ayudara a responder las preguntas de negocio.



Number of dates	holiday	Holiday	17
No Holiday	No Holiday	336	336

Análisis de Correlaciones

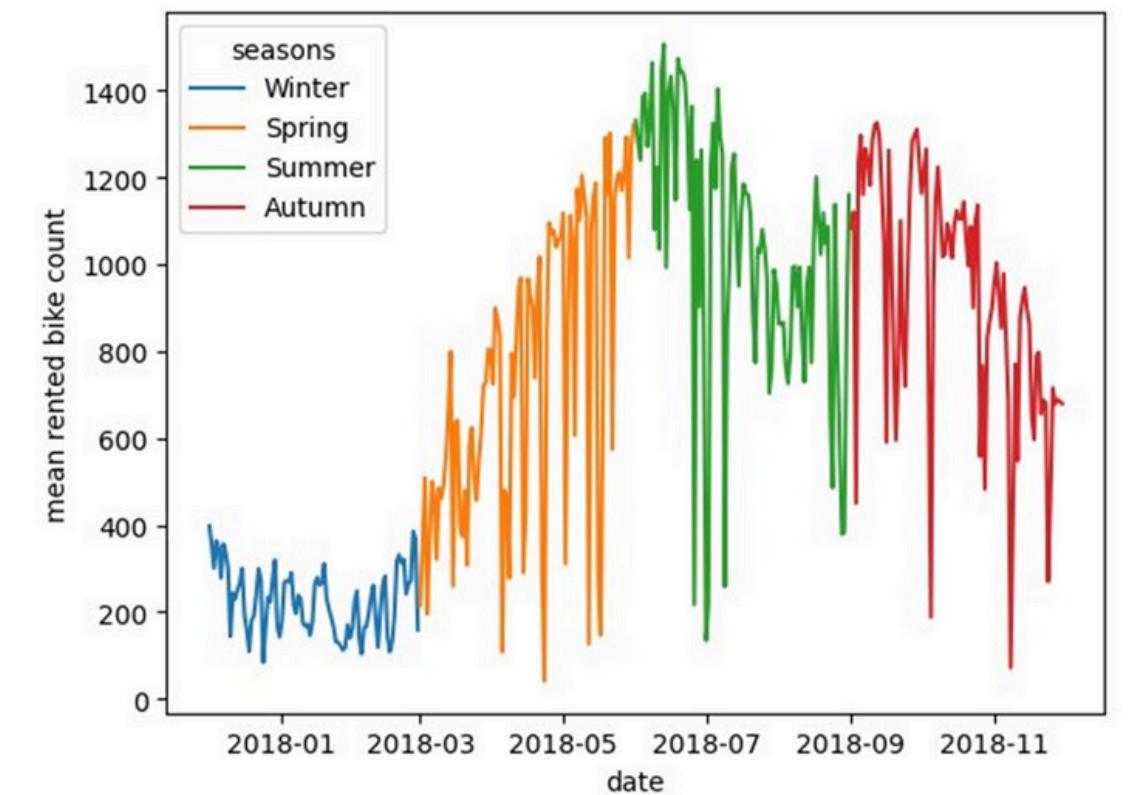
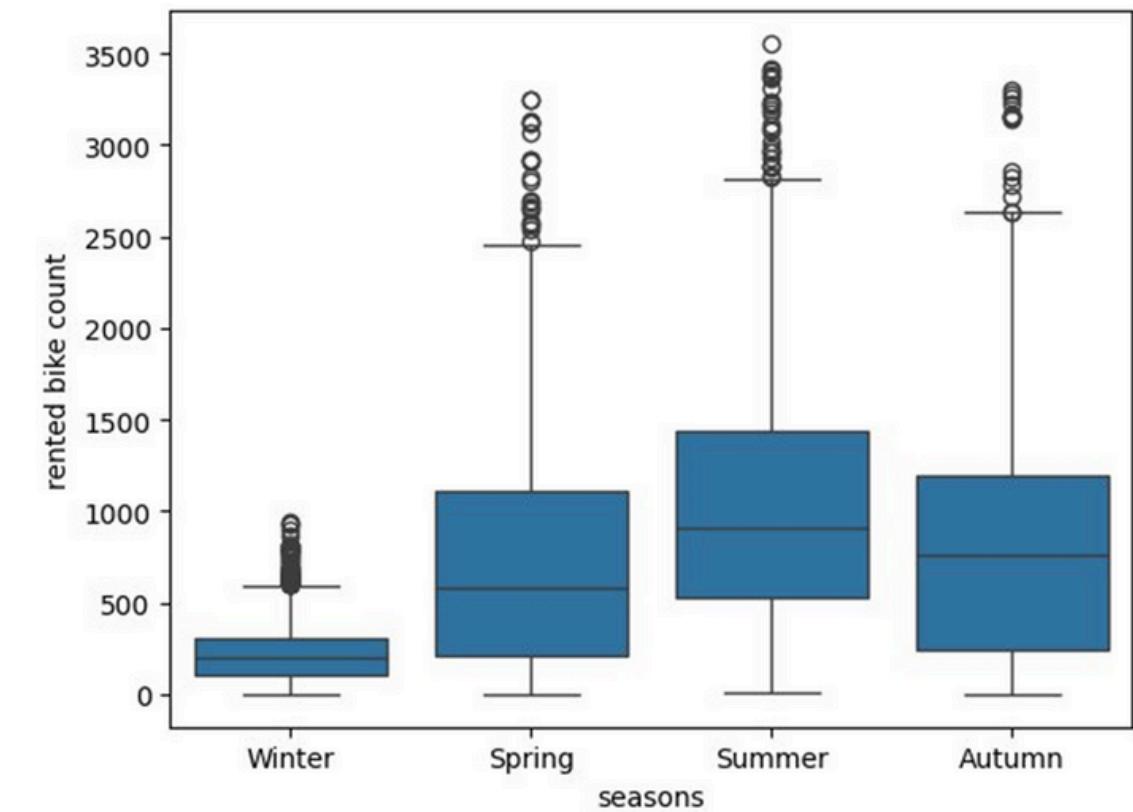


- Observamos una correlación moderada entre la mayoría de variables explicativas, únicamente dew point temperature estaria generando problemas de multicolinealidad.

- En general podemos inferir que a mayor temperatura aumenta el uso de bicicletas rentadas, vemos que esto podría estar explicado por la presencia de estaciones en Seoul, donde en invierno se usa menor cantidad de bicicletas
- Igualmente se puede inferir que en días de mejor condiciones meteorológicas, es decir mayor visibilidad, soleados, con temperatura templada y en mas tarde en el dia el uso de la bicicleta aumentaria

En general, en el análisis de datos se observa:

- Una gran influencia de las variables explicativas restantes sobre la variable de respuesta, por lo cual una regresión lineal le podría ayudar al equipo de operaciones que características climáticas puede generar un mayor uso de las bicicletas en un día determinado.
- Un comportamiento estacional de la cantidad de bicicletas rentadas a lo largo del tiempo, por lo que se recomienda realizar un forecast para prever la cantidad de bicicletas necesarias para el mes de diciembre de 2018



Limpieza Final

- No se encontraron valores perdidos ni entradas duplicadas
- Se transformó el formato de las fechas con Datetime
- Se transformó Seasons en columna numérica

Season	Valor numérico
Autumn	0
Spring	1
Summer	2
Winter	3

Variables características
Date
Hour
Temperature
Humidity
Wind Speed
Visibility
Solar Radiation
Seasons



MODELAMIENTO

Regresión lineal 1

$$y = \beta_0 + \beta_1 \times \text{Hour} + \beta_2 \times \text{Temperature} + \beta_3 \times \text{Humidity} + \beta_4 \times \text{Wind speed} + \beta_5 \times \text{Visibility} + \beta_6 \times \text{Solar Radiation} + \beta_7 \times \text{Seasons}$$

Regresión lineal 2

$$y = \beta_0 + \beta_1 \times \text{Hour} + \beta_2 \times \text{Temperature} + \beta_3 \times \text{Humidity} + \beta_4 \times \text{Visibility} + \beta_5 \times \text{Solar Radiation} + \beta_6 \times \text{Seasons}$$

Best model: ARIMA(5, 0, 0)(0, 0, 0)[0] intercept

MODELO	R^2	MAE	MSE	RMSE
Regresión lineal 1	0.53	332.76	198807.73	445.88
Regresión lineal 2	0.53	332.75	199038.74	446.14
ARIMA	0.83	165.81	69303.06	263.25



DISEÑO DE TABLERO

Demanda de Bicicletas en Seúl

Demanda de Bicicletas por Estación

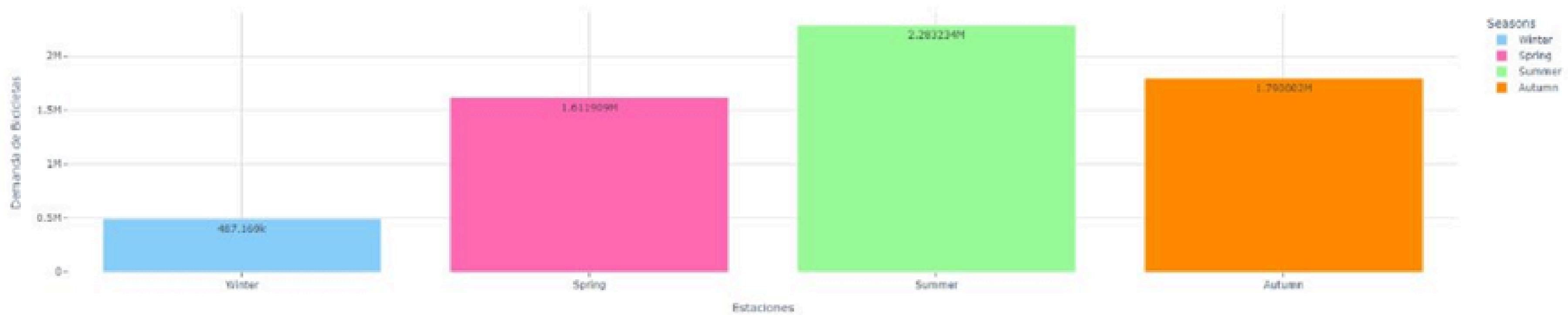


Gráfico de barras

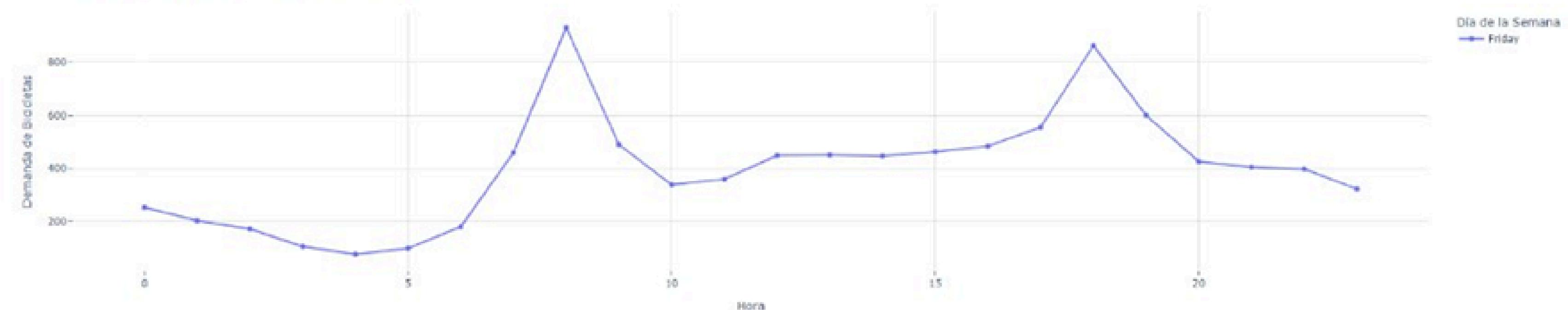
Que relacione la demanda de bicicletas demandadas con las diferentes estaciones del año

Demanda de Bicicletas por Hora

Seleccione una fecha para visualizar la demanda de bicicletas:

2017-12-01

Bicicletas Rentadas por Hora en 2017-12-01



Demanda de Bicicletas por Hora

Seleccione una fecha para visualizar la demanda de bicicletas:

2017-12-01

Bicicletas Rentadas por Hora en 2017-12-01

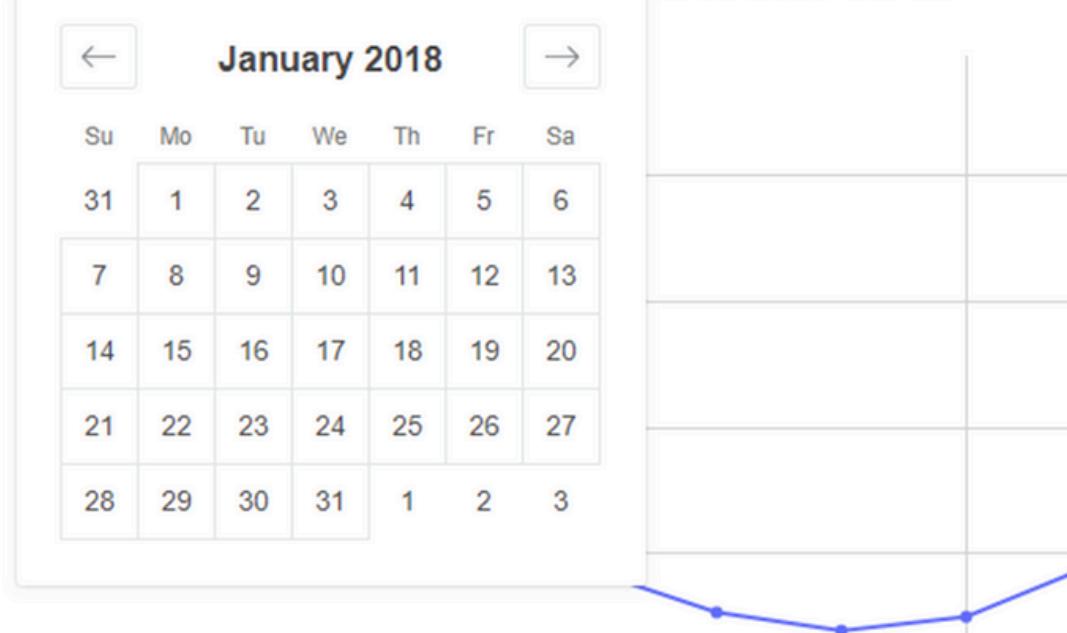
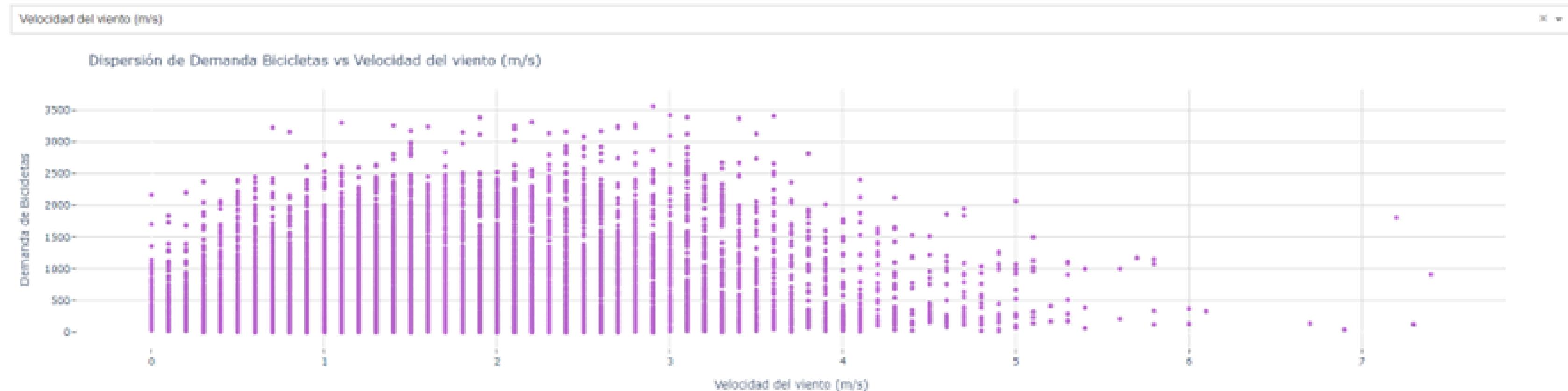


Gráfico de línea

Que muestre la demanda de bicicletas en función de la hora para la fecha seleccionada por el usuario

Demanda de Bicicletas vs. Condiciones Climáticas



Demanda de Bicicletas vs. Condiciones Climáticas

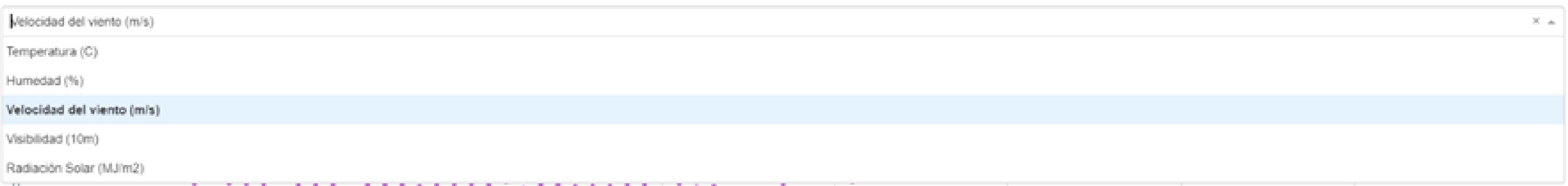


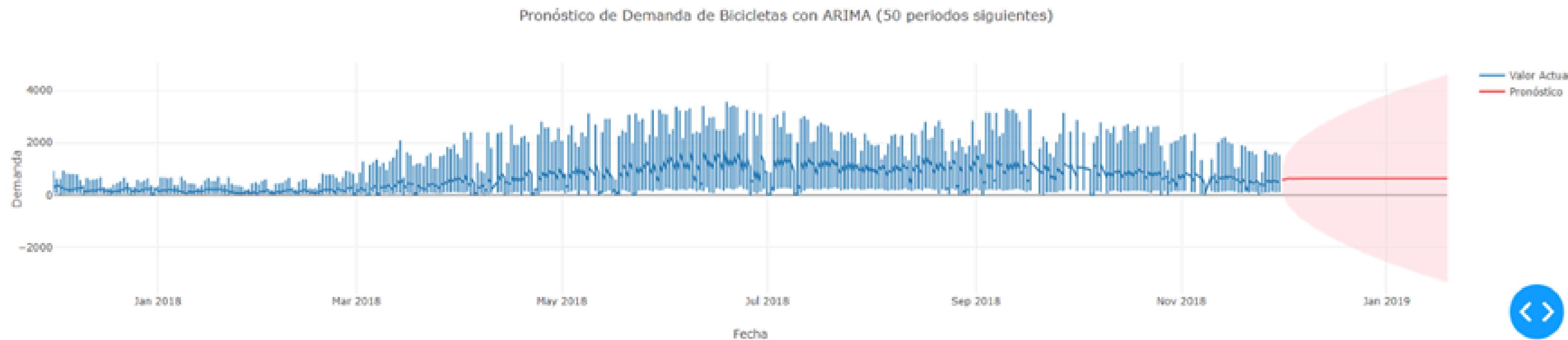
Gráfico de dispersión

Que relacione la demanda de las bicicletas con variables climáticas como la temperatura, humedad, velocidad del viento y radiación solar

Pronóstico de la demanda de bicicletas

para los siguientes 50 períodos a partir de la última fecha

Pronóstico de la Demanda de Bicicletas con ARIMA





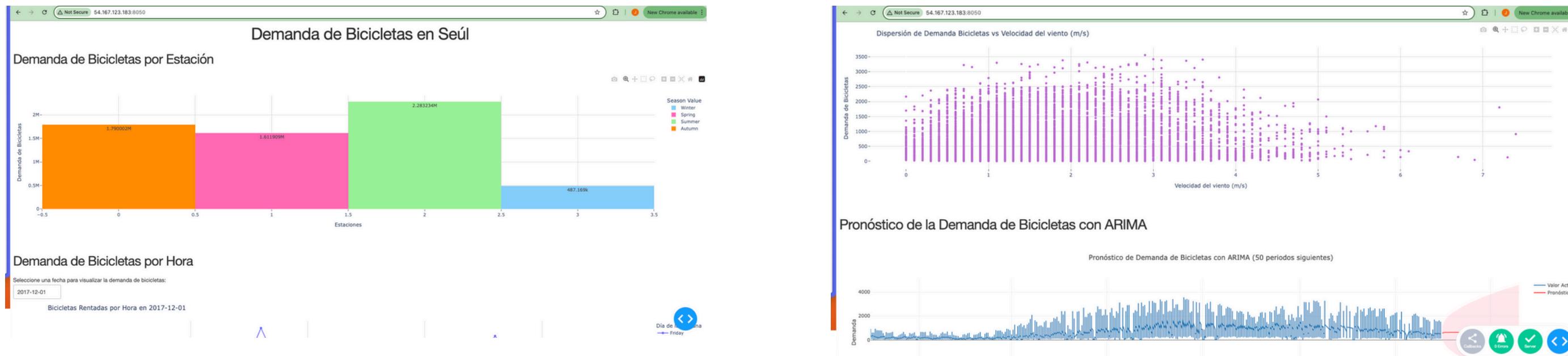
DESPLIEGUE

Despliegue

Características de la máquina virtual utilizada en el despliegue

Instance summary for i-0b11004b64ff1fdf7 (proyecto1_data_trio) Info		
Updated about 2 hours ago		
Instance ID i-0b11004b64ff1fdf7 (proyecto1_data_trio)	Public IPv4 address 54.167.123.183 open address	Private IPv4 addresses 172.31.94.58
IPv6 address -	Instance state Running	Public IPv4 DNS ec2-54-167-123-183.compute-1.amazonaws.com open address
Hostname type IP name: ip-172-31-94-58.ec2.internal	Private IP DNS name (IPv4 only) ip-172-31-94-58.ec2.internal	Elastic IP addresses -
Answer private resource DNS name IPv4 (A)	Instance type t2.micro	AWS Compute Optimizer finding Opt-in to AWS Compute Optimizer for recommendations. Learn more
Auto-assigned IP address 54.167.123.183 [Public IP]	VPC ID vpc-071a45e2884e3528c	Auto Scaling Group name -
IAM Role -	Subnet ID subnet-03fdf65a49dc42f64	
IMDSv2 Required	Instance ARN arn:aws:ec2:us-east-1:471148366428:instance/i-0b11004b64ff1fdf7	

Evidencia del tablero desplegado



<http://54.167.123.183:8050>

CONCLUSIÓN



- La demanda de bicicletas tiene su pico máximo en el verano, sin embargo, siempre se observa que las horas de mayor uso son las 8 am y 6pm.
- Con ayuda de los modelos de regresión lineal se puede concluir que a medida que aumenta la temperatura, la visibilidad y la hora del dia, aumenta la cantidad de ciclas rentadas. Mientras que al aumento de la humedad y radiación se presentaría una caída en la cantidad de ciclas rentadas
- La variable que mayor impacta la demanda de bicicletas rentadas es la temporada del año, según los modelos de regresión lineal y el análisis exploratorio.
- Se espera que el valor promedio de bicicletas rentadas para enero de 2019 sea de 635 bicicletas. Por lo cual pueden utilizar este periodo para hacer mantenimiento de las demás.
- En los datos compartidos se encontraron variables que tienen una alta correlación, y variables que no explican mucho.

**iMUCHAS
GRACIAS!**