

BIKE SHARE

Applied Artificial Intelligence & Machine Learning
AI Data Modeling - Maio 2024

Andreia Gonçalves
Carlos Brás
Carlos Catalão
Pedro Oliveira
Susana Morgado
Thaissa Paz

ÍNDICE



INTRODUÇÃO



DATASET



ANÁLISE EXPLORATÓRIA



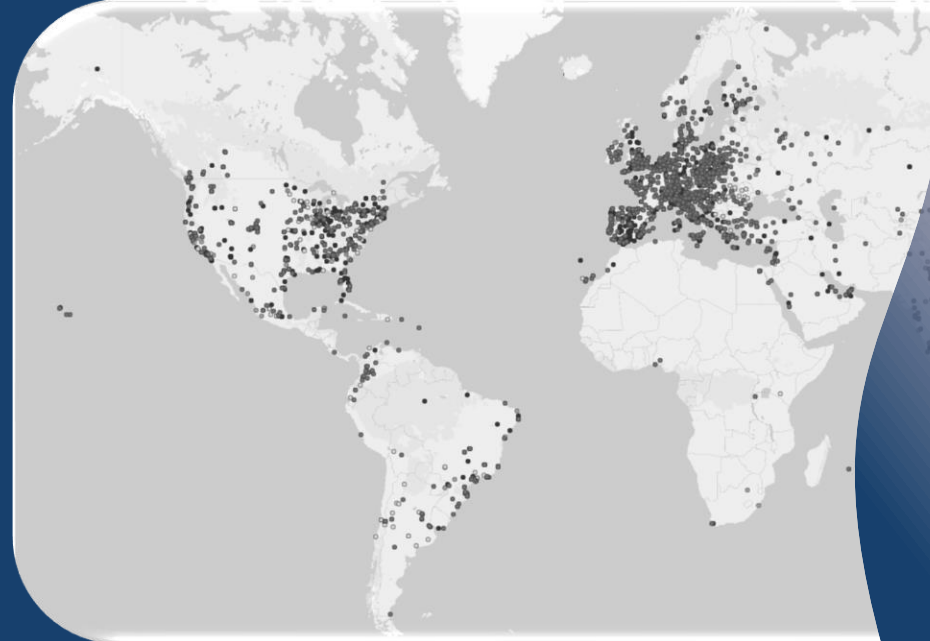
CORRELAÇÃO ENTRE
VARIÁVEIS



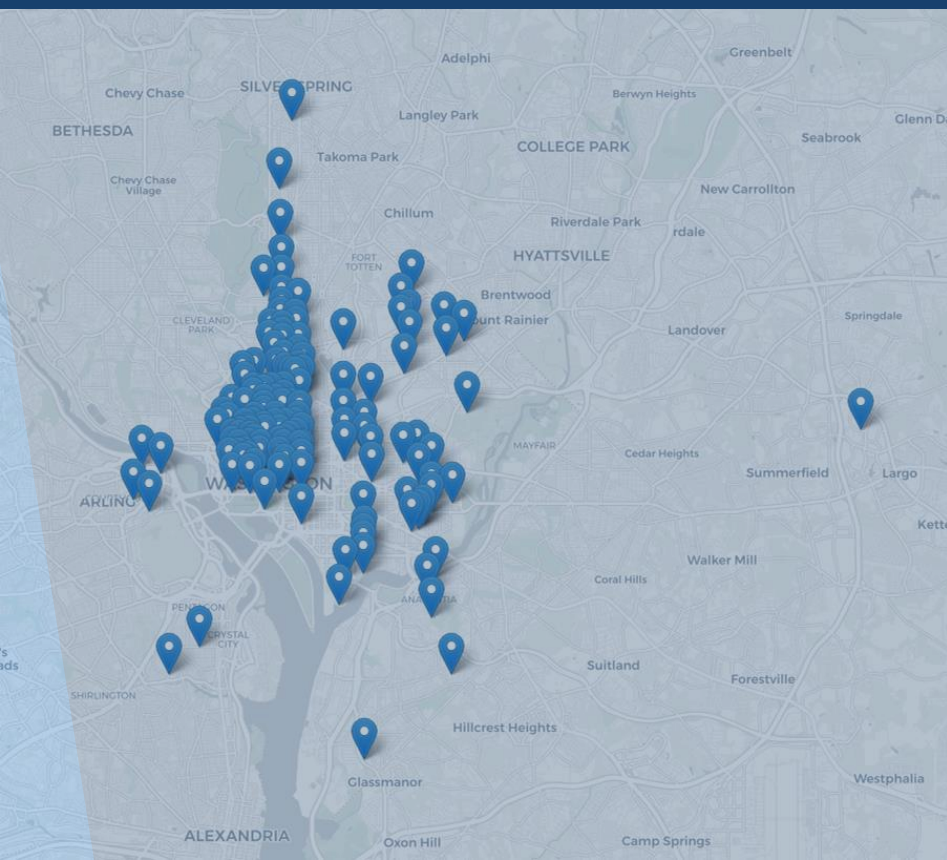
DESENVOLVIMENTOS
FUTUROS

INTRODUÇÃO

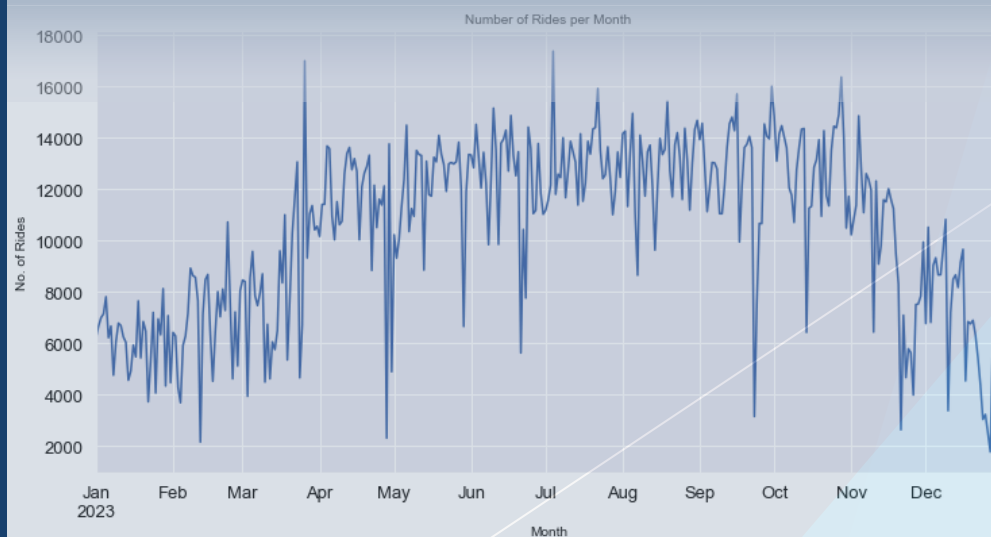
Os sistemas de *Bike Share* permitem que os utilizadores aluguem bicicletas por um valor durante um espaço de tempo, o que possibilita que as pessoas aluguem num determinado ponto e o devolvam noutra ponto diferente, ou que retornem ao ponto de partida. Cada bicicleta pode ser utilizada por diferentes pessoas no mesmo dia.



Nos últimos anos, os sistemas de *Bike Share* têm crescido em todo o mundo



Com este projeto pretendemos analisar os dados de *Bike Share* durante o ano de 2023 na cidade de Washington, nos Estados Unidos da América.





DATASET

Dataset
Bike Share

• 4.467.334 Observações
• 13 Features

Dataset
Weather

• 365 Observações
• 19 Features

TRATAMENTO DE DADOS



Eliminação dos valores omissos
Renomeação de features
Alteração datatype
Criação de novas variáveis
Transformação variáveis em categóricas
Conversão da duração de viagem para segundos

```
# Column
---
0 ride_id
1 rideable_type
2 started_at
3 ended_at
4 start_station_name
5 start_station_id
6 end_station_name
7 end_station_id
8 start_lat
9 start_lng
10 end_lat
11 end_lng
12 member_casual
```

```
# Column
---
0 ride_id
1 rideable_type
2 started_at
3 ended_at
4 start_station_name
5 start_station_id
6 end_station_name
7 end_station_id
8 start_lat
9 start_lng
10 end_lat
11 end_lng
12 user_type
13 started_at_date
14 started_month_name
15 started_weekday
16 started_day
17 started_hour
18 duration_sec
19 season
20 day_type
```

```
ride_id      0.000000
rideable_type 0.000000
started_at   0.000000
ended_at     0.000000
start_station_name 9.102297
start_station_id 9.102297
end_station_name 9.681031
end_station_id 9.681031
start_lat    0.000000
start_lng    0.000000
end_lat      0.133077
end_lng      0.133077
member_casual 0.000000
```

```
ride_id      0.0
rideable_type 0.0
started_at   0.0
ended_at     0.0
start_station_name 0.0
start_station_id 0.0
end_station_name 0.0
end_station_id 0.0
start_lat    0.0
start_lng    0.0
end_lat      0.0
end_lng      0.0
user_type    0.0
```

DATASET FINAL



3 811 987 OBSERVAÇÕES
28 FEATURES

DATASET FINAL

	ride_id	rideable_type	started_at	ended_at	user_type	start_station_name	start_station_id	end_station_name	end_station_id	start_lat	start_lng
0	65F0ACD101BF0D49	classic bike	2023-01-04 19:34:07	2023-01-04 19:39:29	member	East Falls Church Metro / Sycamore St & 19th St N	31904.0	W Columbia St & N Washington St	32609.0	38.885321	-77.156427
1	D75158CE73DC43F0	classic bike	2023-01-27 15:26:38	2023-01-27 19:21:36	member	Carroll & Westmoreland Ave	32025.0	Fenton St & Ellsworth Dr	32036.0	38.975000	-77.011210
2	33E85889625FF7CA	classic bike	2023-01-05 20:44:38	2023-01-05 20:51:18	member	15th & L St NW	31276.0	Thomas Circle	31241.0	38.903649	-77.034918
3	E1F055A1651F47A1	classic bike	2023-01-03 17:45:14	2023-01-03 17:57:23	member	Hartland Rd & Harte Pl	32255.0	Merrifield Cinema & Merrifield Town Center	32235.0	38.878601	-77.222808
4	88CC90CEEC298BAF	classic bike	2023-01-03 05:18:46	2023-01-03 05:25:50	member	Merrifield Cinema & Merrifield Town Center	32235.0	Hartland Rd & Harte Pl	32255.0	38.870093	-77.229970

	end_lat	end_lng	started_at_date	started_month_name	started_weekday	started_day	started_hour	duration_sec	season	day_type	holiday	day_type_all	temp	precip	windspeed	visibility	conditions
0	38.885621	-77.166917	2023-01-04	January	Wednesday	4	19	322	Winter	Weekday	NaN	Weekday	15.3	0.0	22.8	15.9	Partially cloudy
1	38.997033	-77.025608	2023-01-27	January	Friday	27	15	14098	Winter	Weekday	NaN	Weekday	3.9	0.0	23.5	16.0	Partially cloudy
2	38.905900	-77.032500	2023-01-05	January	Thursday	5	20	400	Winter	Weekday	NaN	Weekday	13.5	0.0	20.2	15.9	Partially cloudy
3	38.870093	-77.229970	2023-01-03	January	Tuesday	3	17	729	Winter	Weekday	NaN	Weekday	15.2	0.0	28.5	15.6	Overcast
4	38.878601	-77.222808	2023-01-03	January	Tuesday	3	5	424	Winter	Weekday	NaN	Weekday	15.2	0.0	28.5	15.6	Overcast

	count	duration_sec	start_lat	start_lng	end_lat	end_lng	started_hour	temp	precip	windspeed	visibility
count	3811987.00	3811987.00	3811987.00	3811987.00	3811987.00	3811987.00	3811987.00	3811987.00	3811987.00	3811987.00	3811987.00
mean		1132.79	38.90	-77.03	38.90	-77.03	14.22	17.73	1.66	23.17	15.34
std		5488.24	0.03	0.03	0.03	0.03	4.81	7.39	4.75	7.65	1.58
min		61.00	38.78	-77.37	38.78	-77.37	0.00	-4.20	0.00	9.10	4.80
25%		387.00	38.89	-77.04	38.89	-77.04	11.00	11.90	0.00	17.60	15.50
50%		657.00	38.90	-77.03	38.90	-77.03	15.00	19.00	0.00	21.70	16.00
75%		1139.00	38.91	-77.01	38.91	-77.01	18.00	24.00	0.29	26.00	16.00
max		2939279.00	39.13	-76.83	39.13	-76.83	23.00	30.00	46.04	58.10	16.00

	count	start_station_name	end_station_name	rideable_type	user_type	started_month_name	started_weekday	season	day_type_all	conditions
count		3811987	3811987	3811987	3811987	3811987	3811987	3811987	3811987	3811987
unique		792	797	3	2	12	7	4	3	6
top		columbus circle / union station	columbus circle / union station	classic bike	member	July	Thursday	Summer	Weekday	Partially cloudy
freq		44651	44790	2909059	2362327	402686	577142	1203384	2628108	2171965

VARIÁVEIS
NUMÉRICAS
RELEVANTES

VARIÁVEIS
CATEGÓRICAS
RELEVANTES



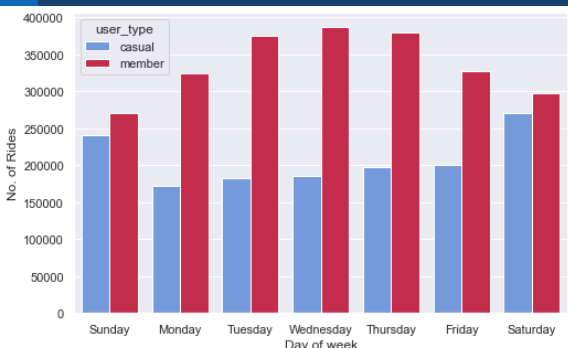
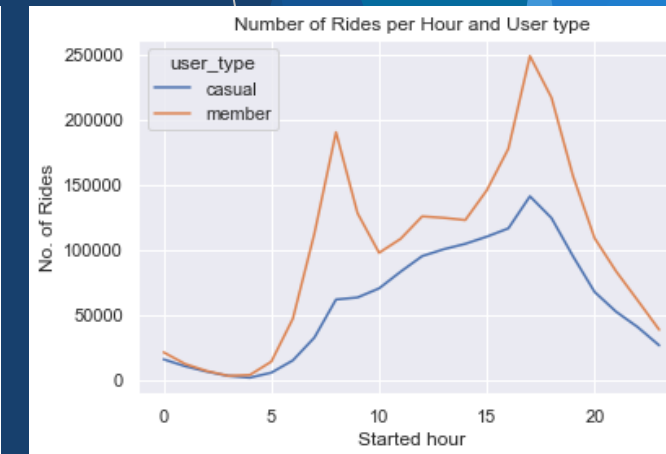
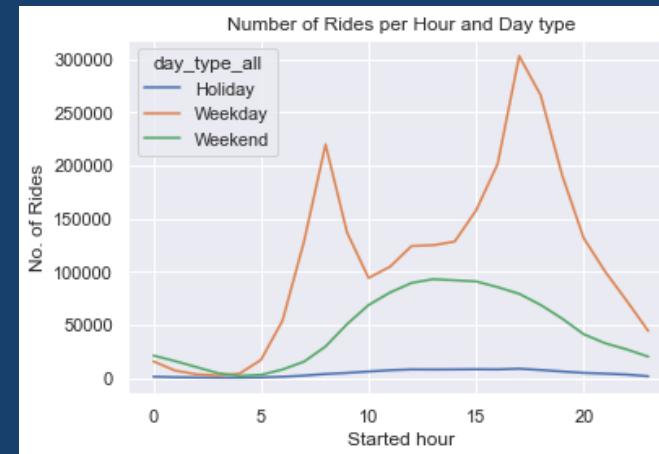
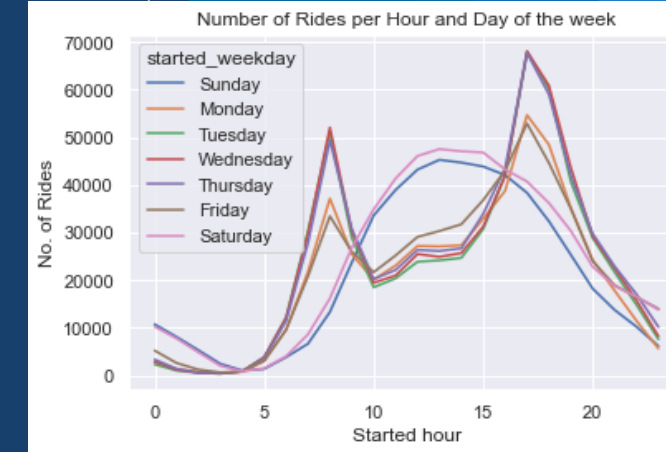
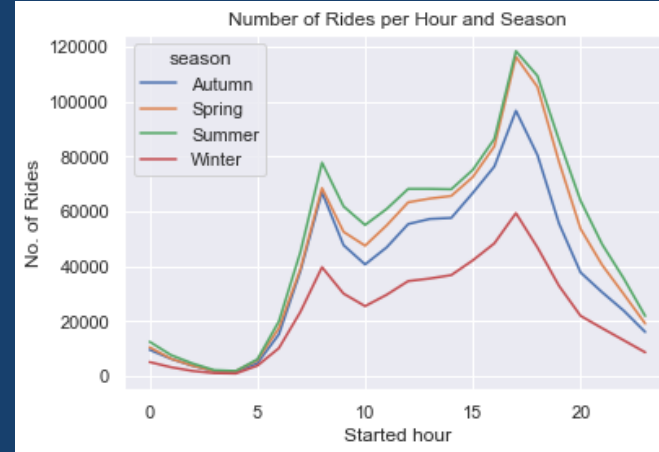
ANÁLISE EXPLORATÓRIA

Como é de esperar, no Inverno há uma redução do número de viagens.

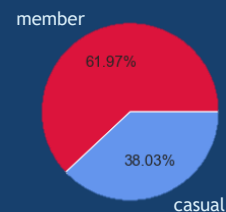
Nos fins de semana, tanto os sábados como os domingos têm a mesma tendência de utilização, atingindo o pico à hora de almoço.

Já durante a semana o padrão é o mesmo para os cinco dias, onde os utilizadores privilegiam a utilização de bicicletas no início e no fim do dia, coincidindo com o início e fim do horário laboral.

É de notar, pelo padrão de utilização, que os utilizadores que são membros são quem utiliza mais bicicletas durante a semana.



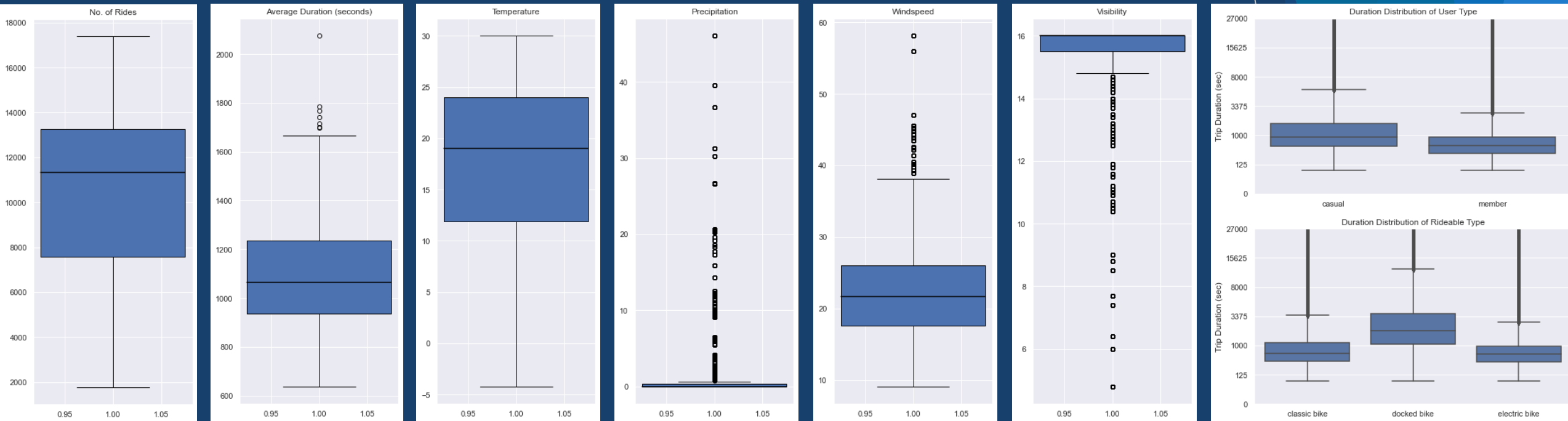
There is more subscribed members than casual users





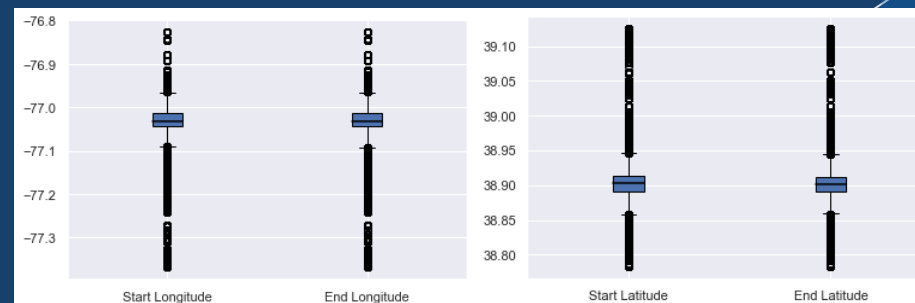
ANÁLISE EXPLORATÓRIA

ANÁLISE DE OUTLIERS



Algumas variáveis apresentam *outliers*.

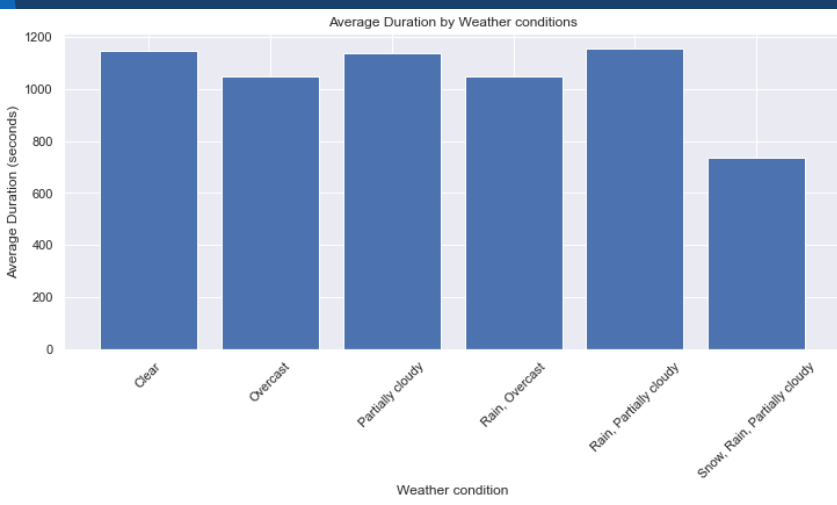
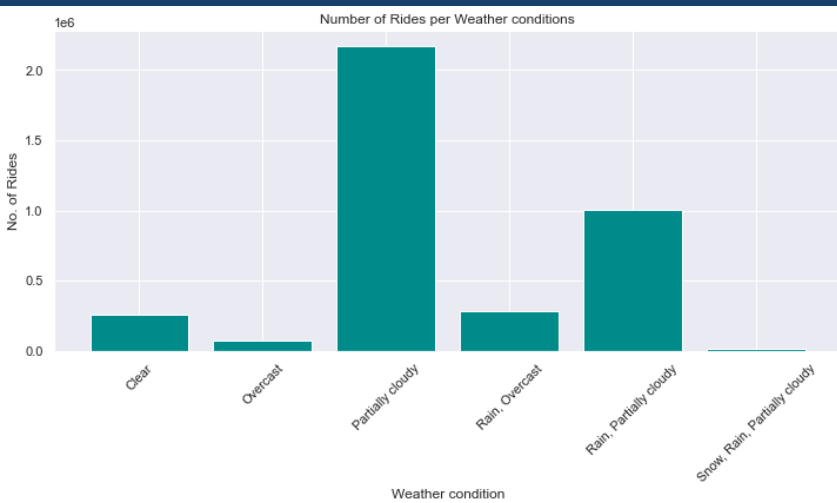
No entanto, por estes aparentarem ser “legítimos” devido às condições atmosféricas (como por exemplo, a tempestade de Agosto de 2023), estes não foram retirados para a análise posterior.



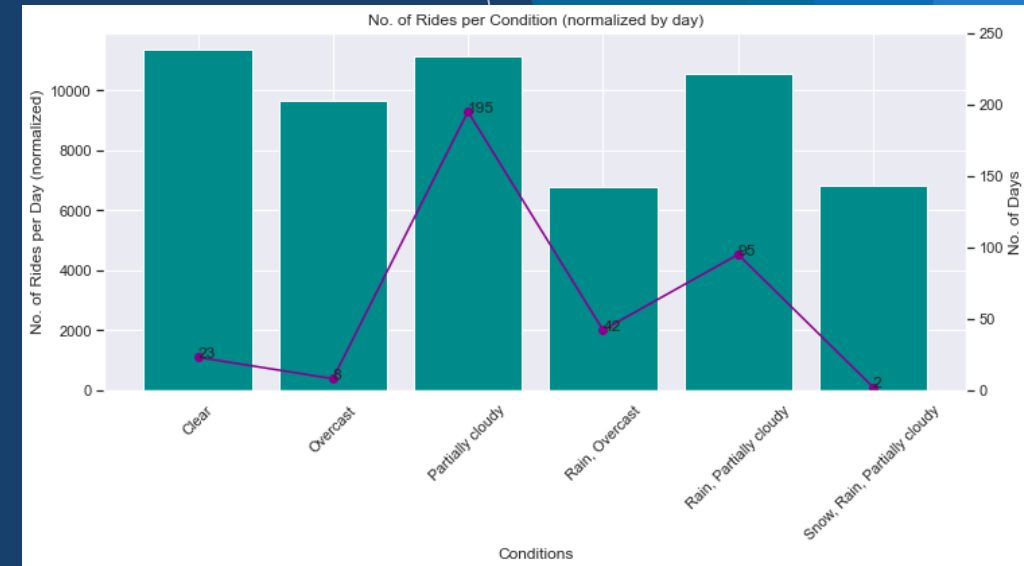


ANÁLISE EXPLORATÓRIA

À primeira vista, os utilizadores parecem ter preferência por dias parcialmente nublados.



No entanto, ao comparar a frequência de utilizações por dia para cada condição meteorológica, é possível identificar que, embora *Partially cloudy* tenha o maior número total de utilizações (devido aos 195 dias num ano), a sua frequência de utilizações por dia é comparável às restantes condições meteorológicas.

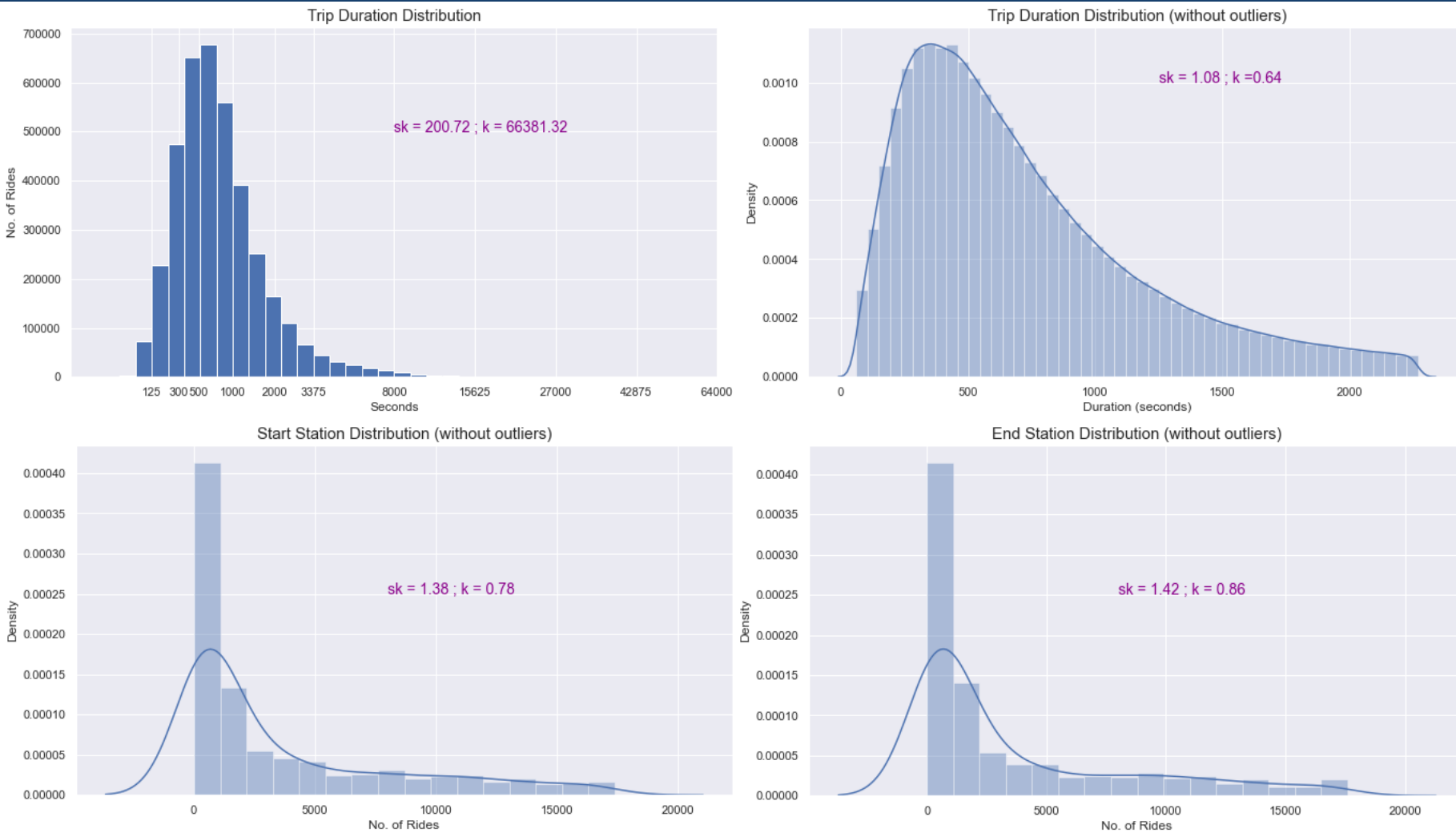


Assim, verifica-se uma análise mais precisa da relação entre as utilizações e as condições, evitando interpretações enviesadas pelo número bruto de dias em cada categoria.

Em termos de duração média das viagens, não há grandes diferenças para os diferentes estados do tempo.



ANÁLISE EXPLORATÓRIA



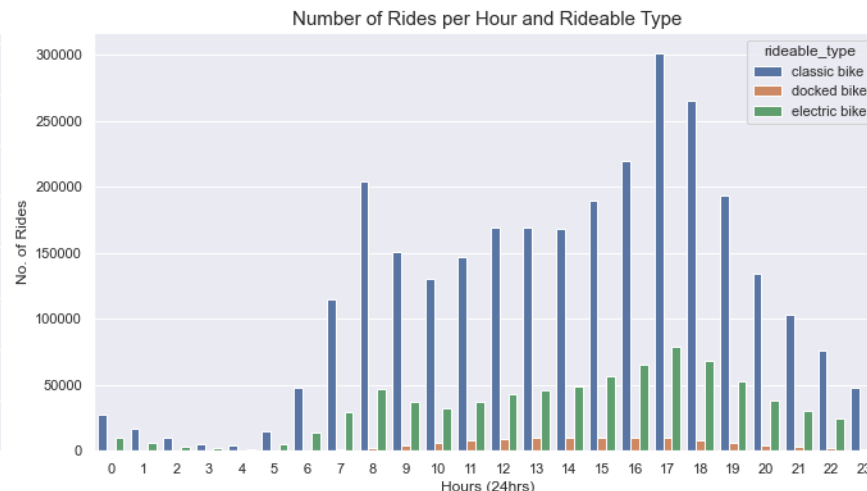
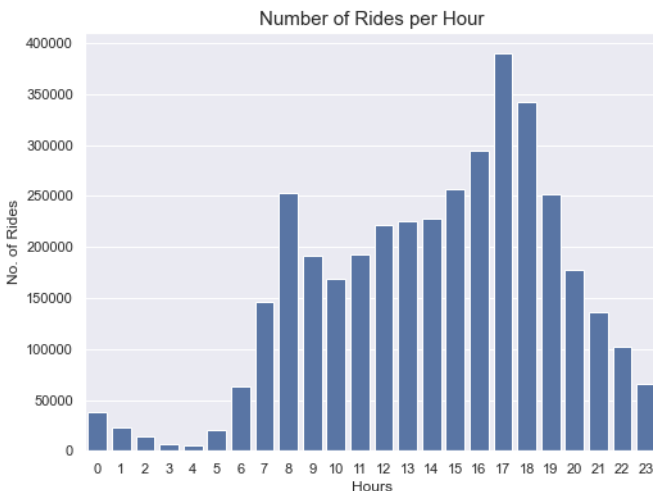
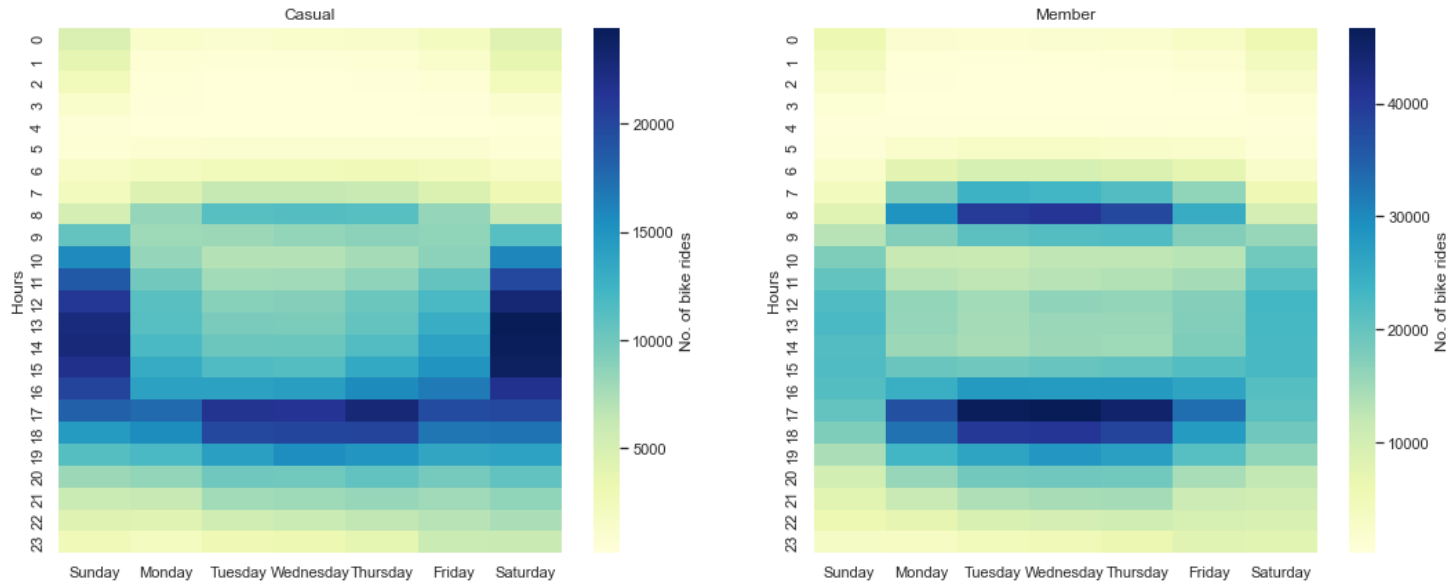
A distribuição da Duração das Viagens tem assimetria positiva, indicando que a média é inferior à mediana e a moda é superior à média e mediana.

O coeficiente de *kurtosis* é bastante elevado, devido a uma maior concentração de valores extremos ou discrepantes nas caudas.

Tanto a distribuição das Start Station como a das End Station têm assimetria positiva e, apesar de leptocúrticas, o coeficiente de *kurtosis* é muito próximo de zero.



ANÁLISE EXPLORATÓRIA



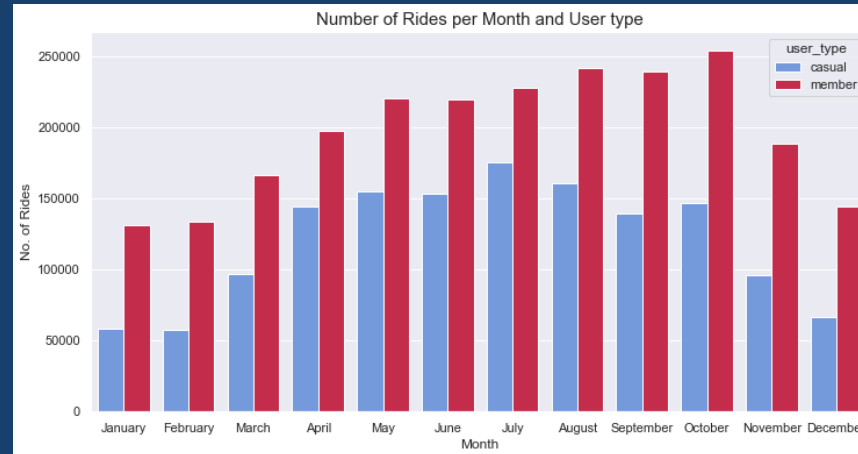
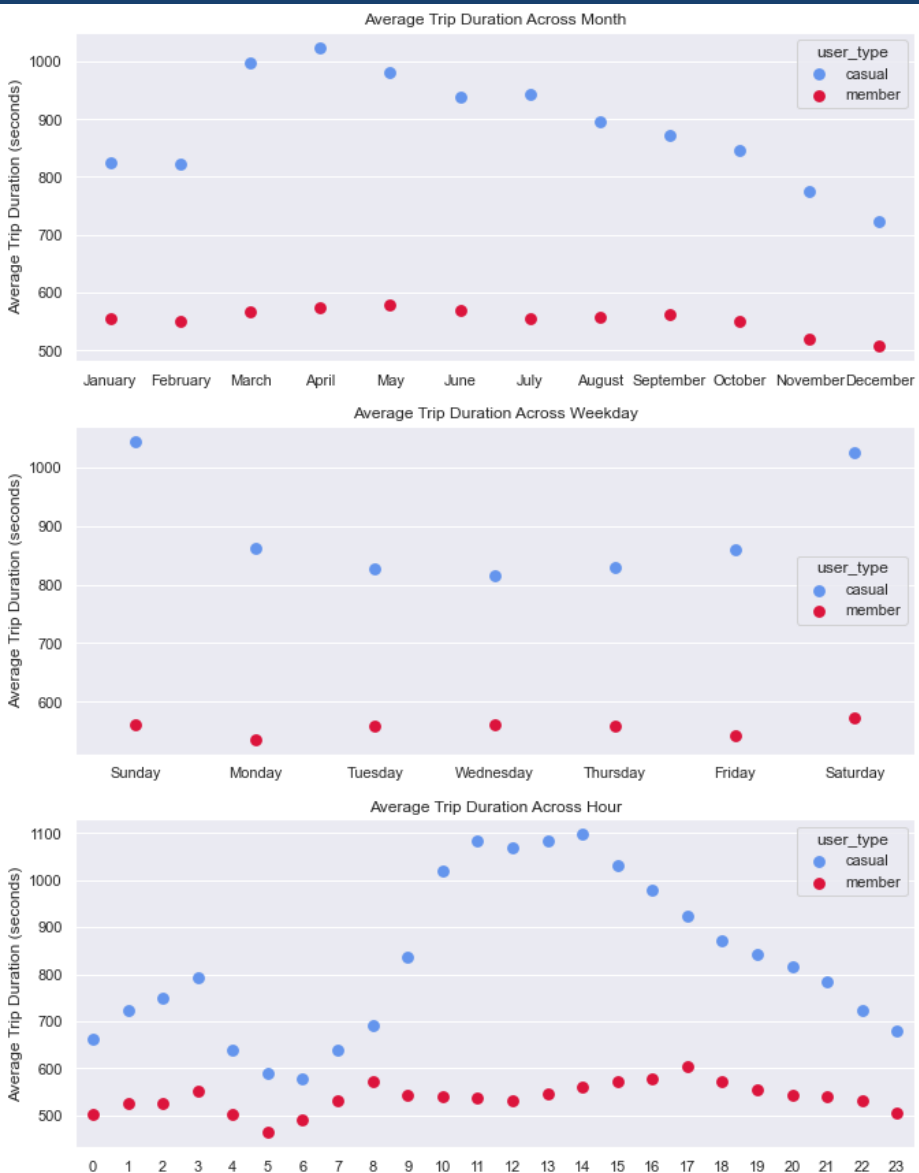
Analisando a utilização de bicicletas por horas, verifica-se que existe uma maior frequência de viagens durante a semana no início e no fim do dia (8h e 17h, respetivamente), coincidindo com o início e fim do horário laboral.

Já ao fim-de-semana os utilizadores preferem aproveitar o dia a andar de bicicleta durante o início da tarde.

Este padrão é idêntico em qualquer tipo de bicicleta utilizado (clássico, *docked* ou elétrico).

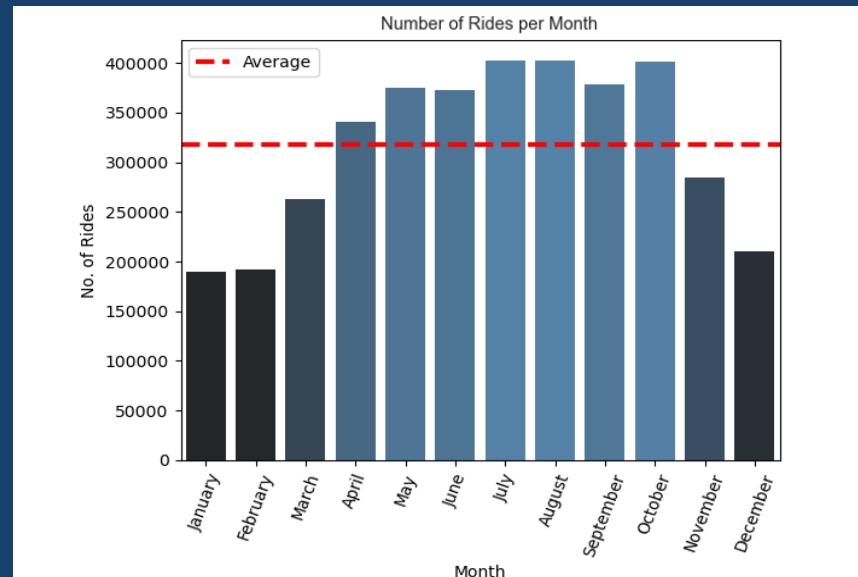


ANÁLISE EXPLORATÓRIA



A duração média das viagens são consistentes ao longo dos meses e dias da semana para utilizadores membros, o que corrobora a teoria de que estes utilizadores regulares usam a bicicleta para ir trabalhar.

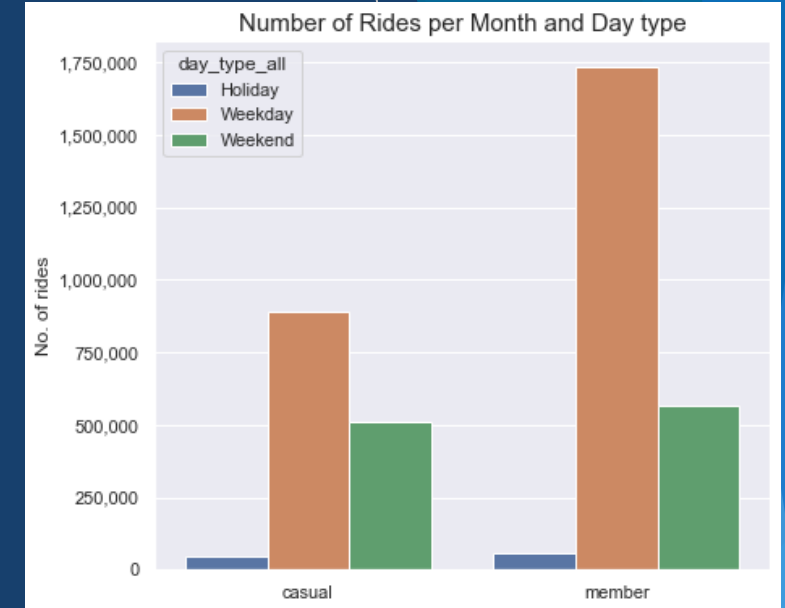
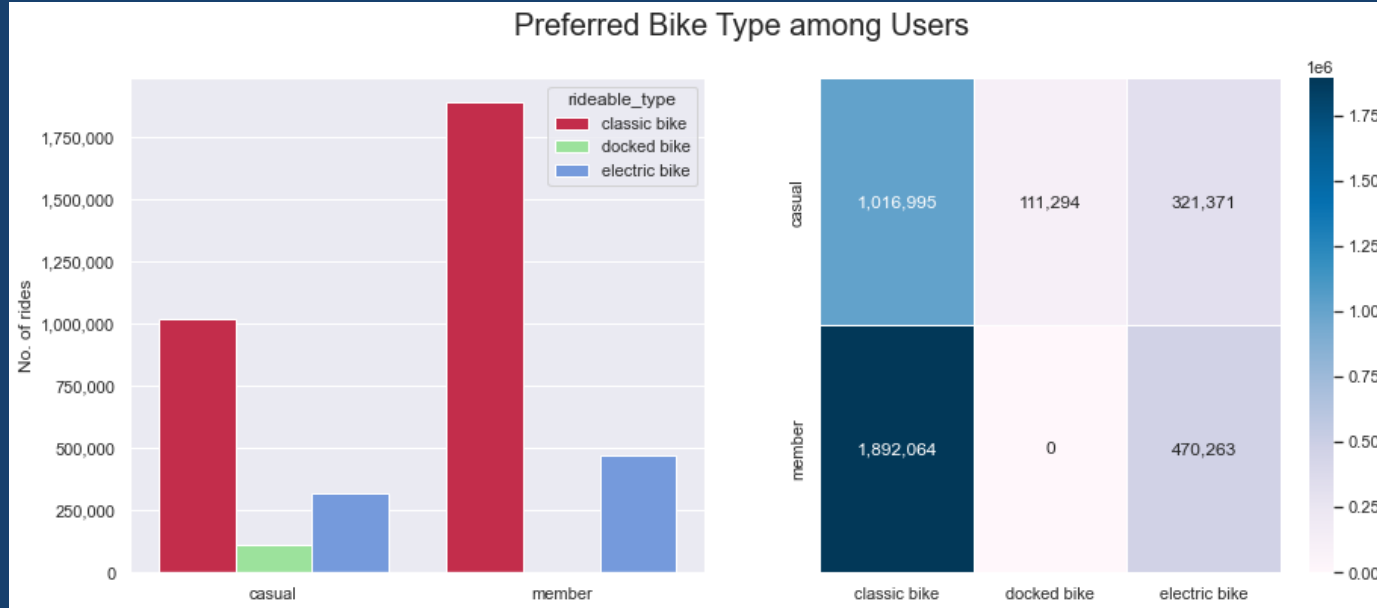
Para os utilizadores casuais existem variações na duração média das viagens, o que nos mostra que poderão ser, na sua maioria, turistas que andam de bicicleta durante mais tempo aos fins-de-semana de manhã na Primavera.



Em termos de número total de viagens, destacam-se os meses de Abril a Outubro, tanto para membros casuais como para membros.



ANÁLISE EXPLORATÓRIA



Tanto os utilizadores casuais como os membros têm preferência pelas bicicletas clássicas, talvez pela possibilidade do aluguer ser mais caro para bicicletas elétricas.

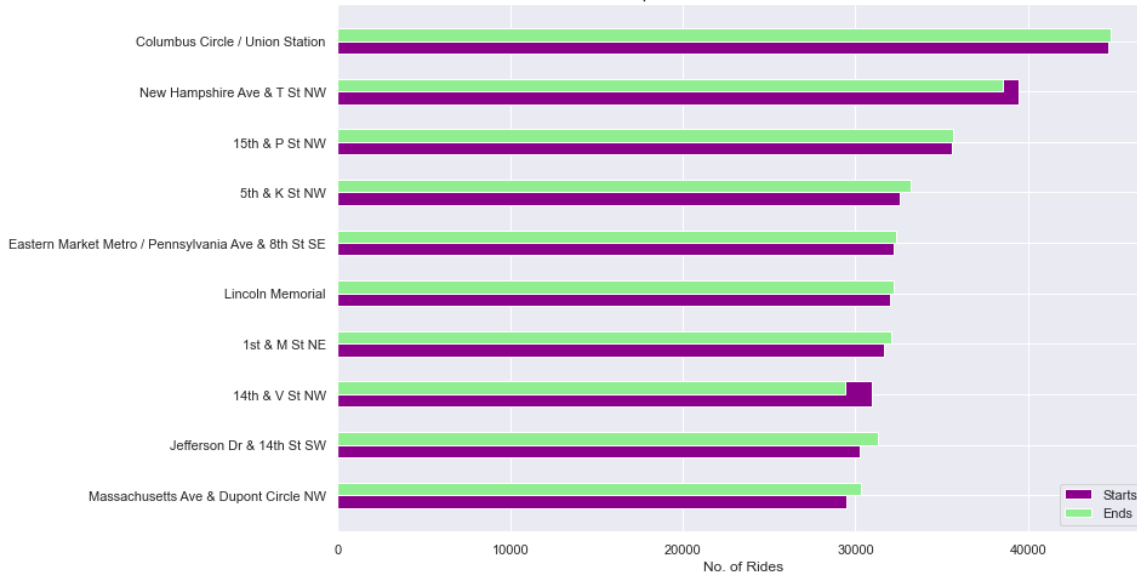
Durante a semana há mais alugueres de bicicletas por parte de membros. Já durante o fim-de-semana não há uma diferença significativa entre o tipo de utilizador.

Como apenas temos 12 feriados no ano de 2023 em Washington, a informação para estes dias é pouco relevante.

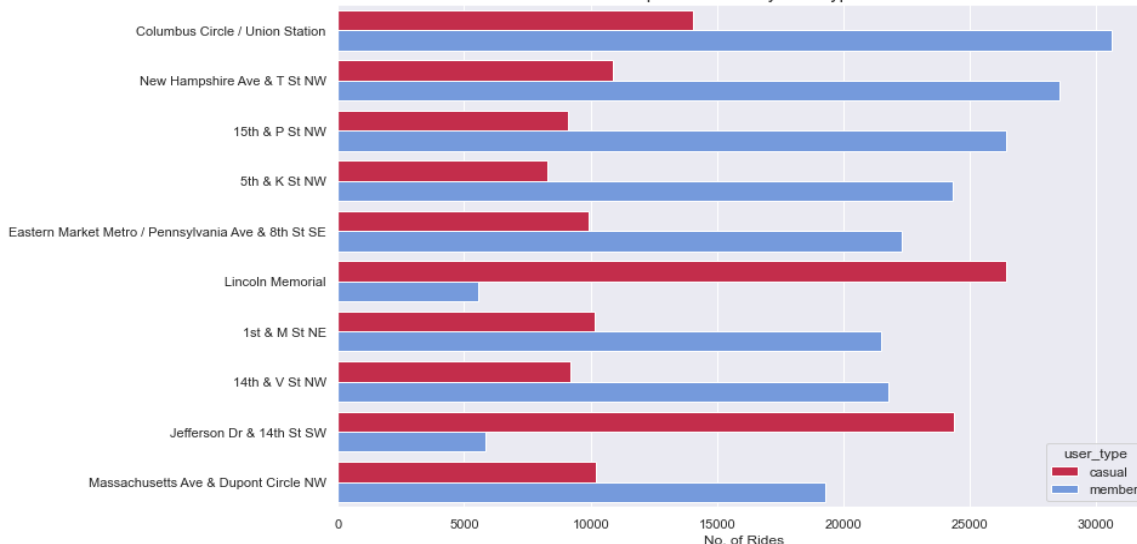


ANÁLISE EXPLORATÓRIA

Top 10 Start and End Stations



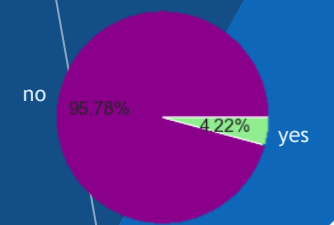
Top 10 Stations by User type



As 10 estações mais utilizadas no início da viagem são também as mais utilizadas para terminar a viagem.

No entanto, em apenas 4% das viagens a estação de chegada é a mesma que a estação de partida.

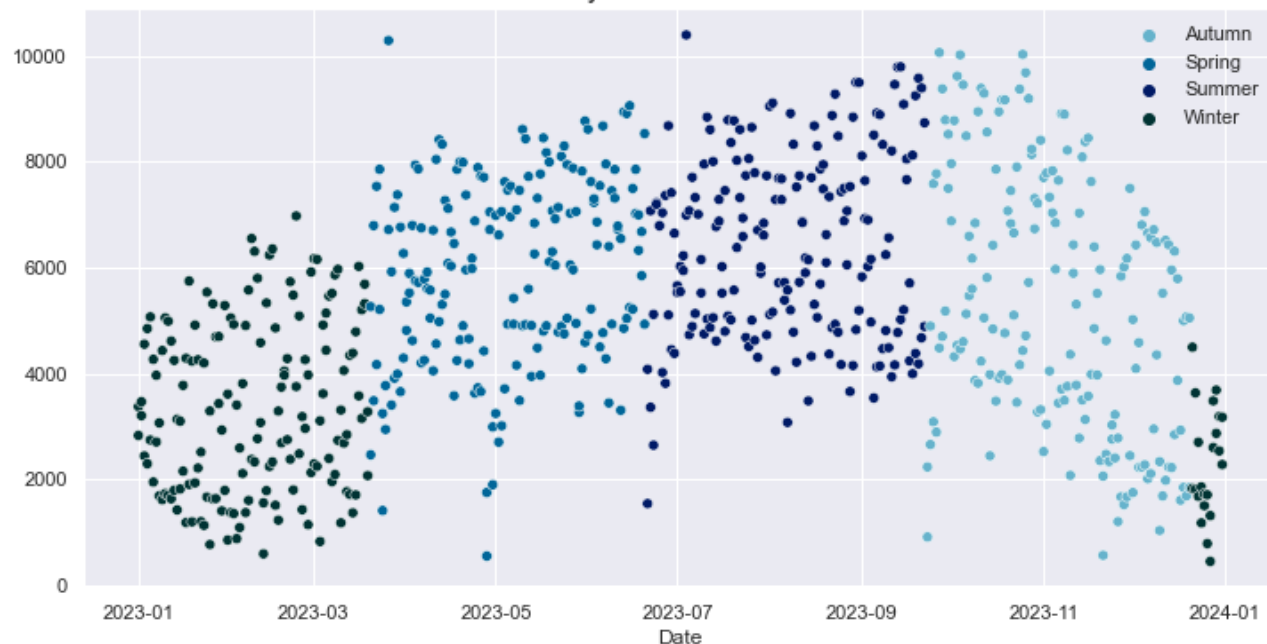
Is the end station the same as the start station?



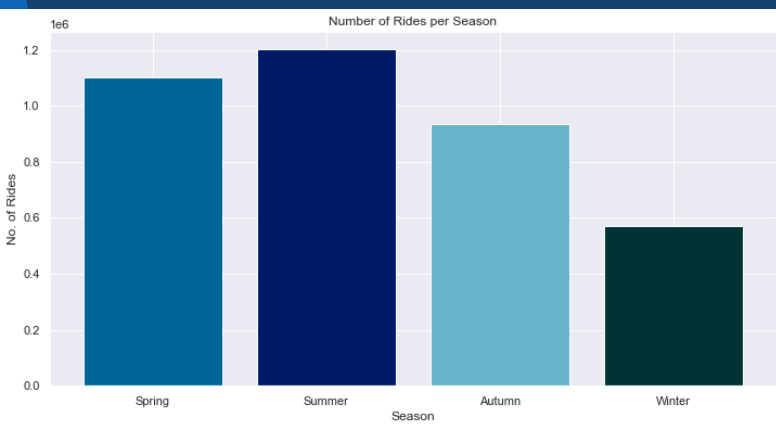
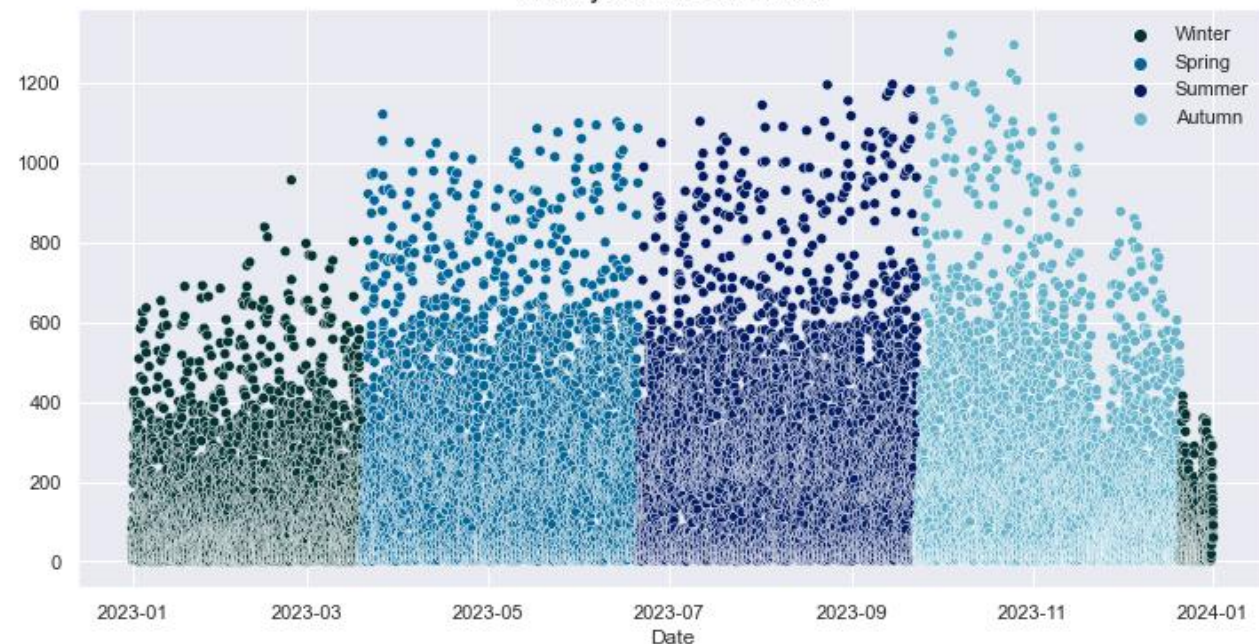


CORRELAÇÃO ENTRE VARIÁVEIS

Daily Number of Rides



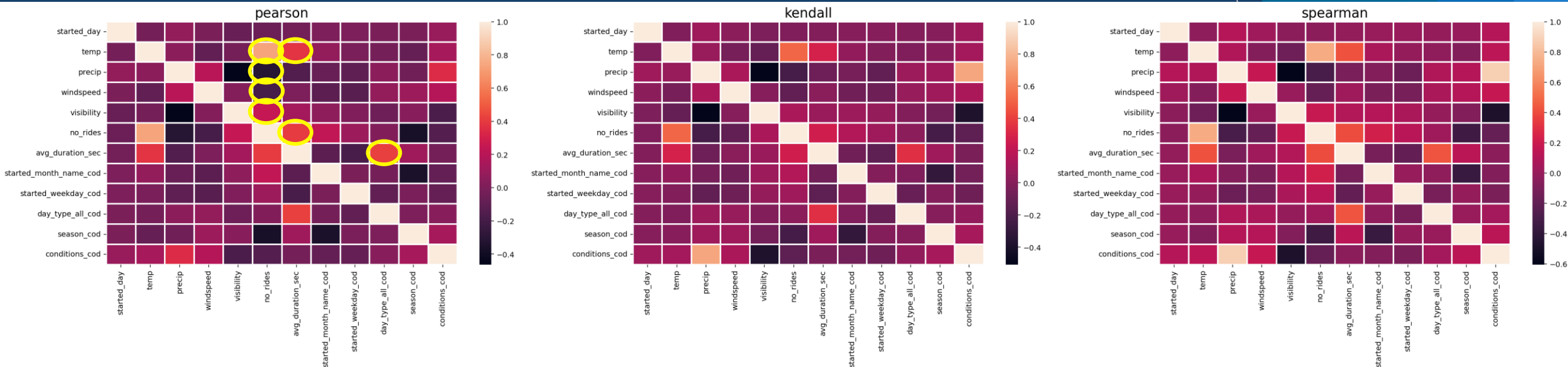
Hourly Number of Rides



Houve mais viagens nos meses de Verão, seguido da Primavera, aparentando uma tendência sazonal.



CORRELAÇÃO ENTRE VARIÁVEIS

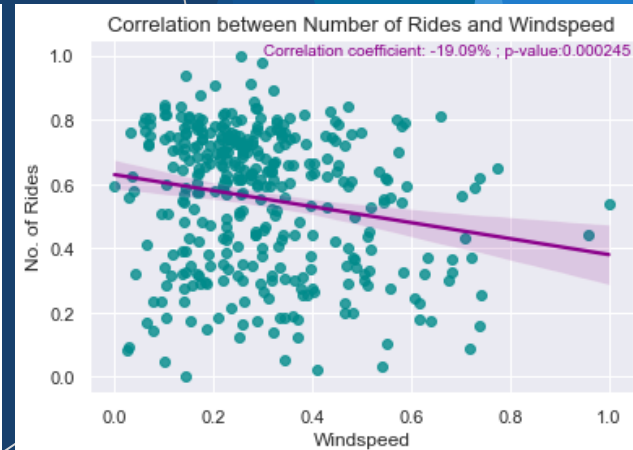
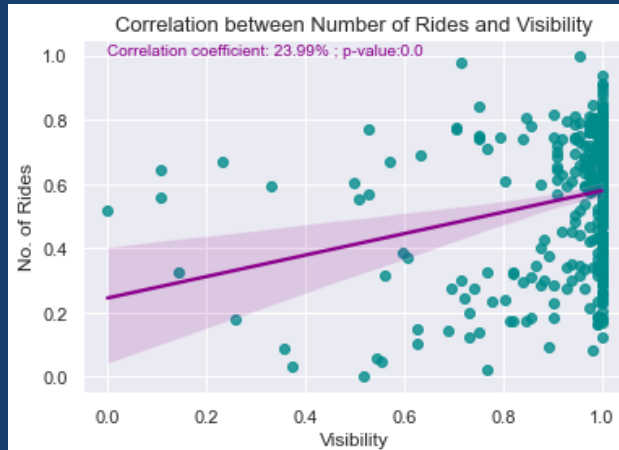
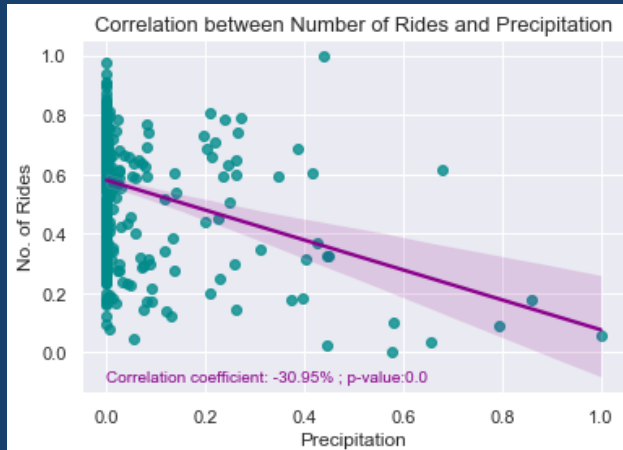
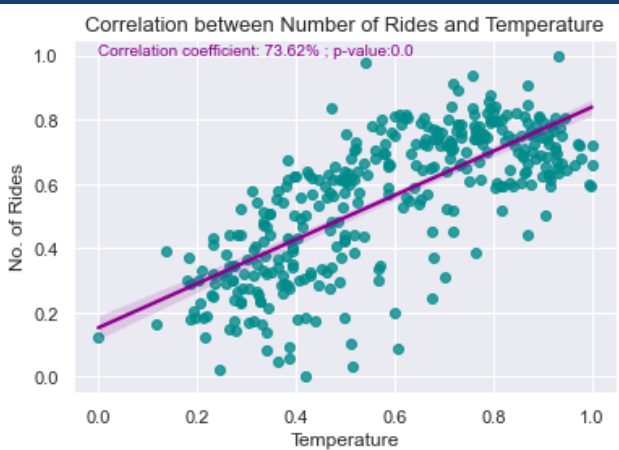
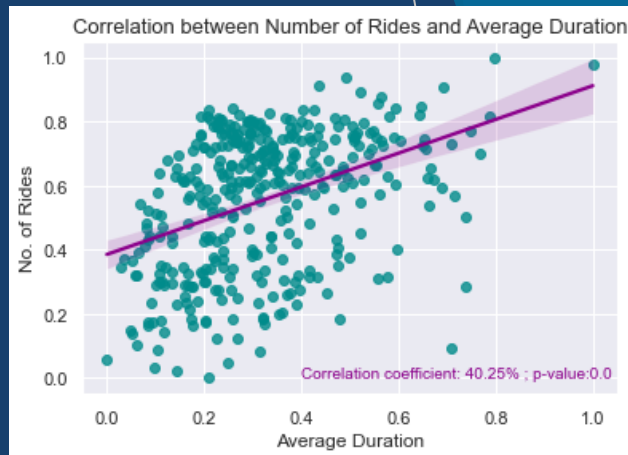
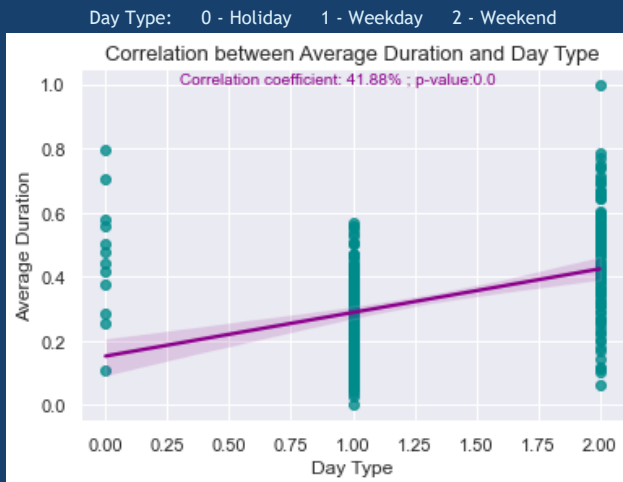
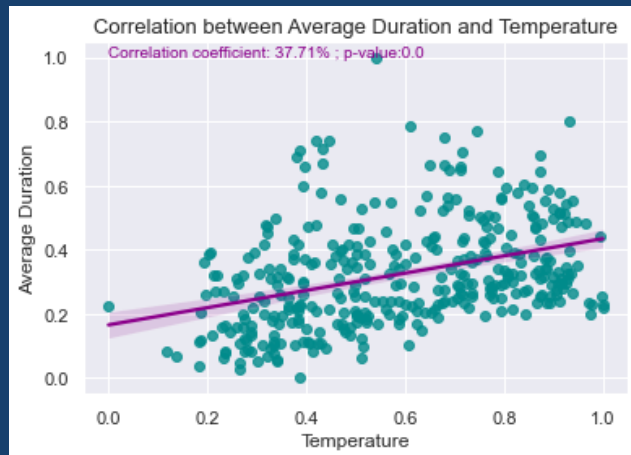


Algumas correlações interessantes para análise individual

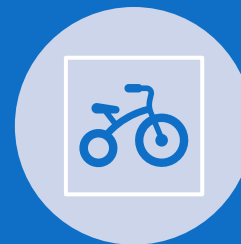
Variáveis		Coefficiente de Correlação
Average Duration	Temperature	37.71%
Average Duration	Day Type	41.88%
Number of Rides	Average Duration	40.25%
Number of Rides	Temperature	73.62%
Number of Rides	Precipitation	-30.95%
Number of Ride	Visibility	23.99%
Number of Rides	Windspeed	-19.09%



CORRELAÇÃO ENTRE VARIÁVEIS



DESENVOLVIMENTOS FUTUROS



PREVISÃO DO NÚMERO DE
UTILIZAÇÕES NOS PRÓXIMOS
DIAS COM BASE NUM MODELO
DE SÉRIES TEMPORAIS



AJUSTAR A OFERTA
ANTECIPANDO O NÚMERO DE
BICICLETAS UTILIZADAS EM
CADA ESTAÇÃO DE FORMA A
COBRIR A PROCURA



DELINEAR ESTRATÉGIAS COM
O OBJETIVO DE CONVERTER
UTILIZADORES CASUAIS EM
MEMBROS

OBRIGADO

Applied Artificial Intelligence & Machine Learning
AI Data Modeling - Maio 2024

Andreia Gonçalves
Carlos Brás
Carlos Catalão
Pedro Oliveira
Susana Morgado
Thaissa Paz