

Atlântico Academy bootcamp



ANÁLISE EXPLORATORIA DE DADOS

Sistema de bicicletas pública

Vitor Costa
Deigela Lima
Hugo Santos
Thais Faustino

Professor: Madson Dias
Monitor(a): Thayná Fuzia

SUMÁRIO:

1 Descrição dos dados

- Configurações iniciais
- Dicionário de dados
- Dados Faltantes

2 Perguntas de partida e hipótese

3 Insights

4 Conclusão



1 Descrição dos dados

■ Configurações iniciais

```
from IPython.display import Markdown
import pandas as pd
import seaborn as sns
```

```
df_hour = pd.read_csv('hour.csv')
display(Markdown("### Rota_Bikers - Hour dataset"))
display(df_hour.head())
```

```
display(Markdown("### Analise exploratória de dados"))
df_hour.info()
```

```
display(Markdown("### Estatísticas descritivas"))
display(df_hour.describe())
```

Análise exploratória de dados

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 731 entries, 0 to 730
Data columns (total 16 columns):
#   Column      Non-Null Count  Dtype
---  -
0   instant     731 non-null   int64
1   dteday      731 non-null   object
2   season      731 non-null   int64
3   yr          731 non-null   int64
4   mnth        731 non-null   int64
5   holiday     731 non-null   int64
6   weekday     731 non-null   int64
7   workingday  731 non-null   int64
8   weathersit   731 non-null   int64
9   temp        731 non-null   float64
10  atemp       731 non-null   float64
11  hum         731 non-null   float64
12  windspeed   731 non-null   float64
13  casual      731 non-null   int64
14  registered  731 non-null   int64
15  cnt         731 non-null   int64
dtypes: float64(4), int64(11), object(1)
memory usage: 91.5+ KB
```



1 Descrição dos dados

■ Configurações iniciais

Rota_Bikers - Day dataset

	instant	dteday	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	hum	windspeed	casual	i
0	1	2011-01-01	1	0	1	0	6	0	2	0.344167	0.363625	0.805833	0.160446	331	
1	2	2011-01-02	1	0	1	0	0	0	2	0.363478	0.353739	0.696087	0.248539	131	
2	3	2011-01-03	1	0	1	0	1	1	1	0.196364	0.189405	0.437273	0.248309	120	
3	4	2011-01-04	1	0	1	0	2	1	1	0.200000	0.212122	0.590435	0.160296	108	
4	5	2011-01-05	1	0	1	0	3	1	1	0.226957	0.229270	0.436957	0.186900	82	



1 Descrição dos dados



■ Configurações iniciais

Estatísticas descritivas

	instant	season	yr	mnth	holiday	weekday	workingday	weathersit	temp	atemp	registered
count	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000	731.000000
mean	366.000000	2.496580	0.500684	6.519836	0.028728	2.997264	0.683995	1.395349	0.495385	0.474354	0.499814
std	211.165812	1.110807	0.500342	3.451913	0.167155	2.004787	0.465233	0.544894	0.183051	0.162961	0.177556
min	1.000000	1.000000	0.000000	1.000000	0.000000	0.000000	0.000000	1.000000	0.059130	0.079070	0.000000
25%	183.500000	2.000000	0.000000	4.000000	0.000000	1.000000	0.000000	1.000000	0.337083	0.337842	0.000000
50%	366.000000	3.000000	1.000000	7.000000	0.000000	3.000000	1.000000	1.000000	0.498333	0.486733	0.000000
75%	548.500000	3.000000	1.000000	10.000000	0.000000	5.000000	1.000000	2.000000	0.655417	0.608602	0.000000
max	731.000000	4.000000	1.000000	12.000000	1.000000	6.000000	1.000000	3.000000	0.861667	0.840896	0.000000

1 Descrição dos dados

■ Dicionário de dados

	variavel	descricao	tipo	subtipo
0	Instant	Índice de registro	Quantitativa	Discreta
1	Dteday	Data	Qualitativa	Ordinal
2	Season	Estação (1:inverno, 2:primavera, 3:verão, 4:ou...	Qualitativa	Nominal
3	Yr	Ano	Qualitativa	Ordinal
4	Mnth	Mês	Qualitativa	Ordinal
5	Holiday	Se é feriado (1: positivo, 0: negativo)	Qualitativa	Nominal
6	Weekday	Dia da semana	Qualitativa	Ordinal
7	Workingday	Se é dia útil (1: positivo, 0: negativo)	Qualitativa	Nominal
8	Weathersit	1: Céu limpo, 2: Nublado, 3: Chuva fraca, 4: C...	Qualitativa	Nominal
9	Temp	Temperatura normalizada em Celsius. Os valores...	Quantitativa	Contínua
10	Atemp	Temperatura normalizada em Celsius.	Quantitativa	Contínua
11	Hum	Umidade normalizada. Os valores são divididos ...	Quantitativa	Contínua
12	Windspeed	Velocidade do vento normalizada.	Quantitativa	Contínua
13	Casual	Contagem de usuários casuais	Quantitativa	Discreta
14	Registered	Contagem de usuários registrados	Quantitativa	Discreta
15	Cnt	Contagem do total de bicicletas alugadas, incl...	Quantitativa	Discreta

```
df_dict.to_csv('/content/sample_data/dicionario_dados_day.csv', index=False)
```

■ Dados faltantes

```
instant      0
dteday       0
season       0
yr           0
mnth         0
holiday      0
weekday      0
workingday   0
weathersit    0
temp         0
atemp        0
hum          0
windspeed    0
casual       0
registered   0
cnt          0
dtype: int64
```



2 Perguntas de partida e hipótese



- Há diferença entre a quantidade de bicicletas alugadas nos dias úteis e não úteis?
- Existe correlação da condição climática com a quantidade de bicicletas alugadas?
- As estações do ano podem influenciar na quantidade de bicicletas alugadas?
- A quantidade de usuários casuais e registrados oscila ao longo do tempo?

3 Insights



4 Conclusão



Obrigado pela atenção!