

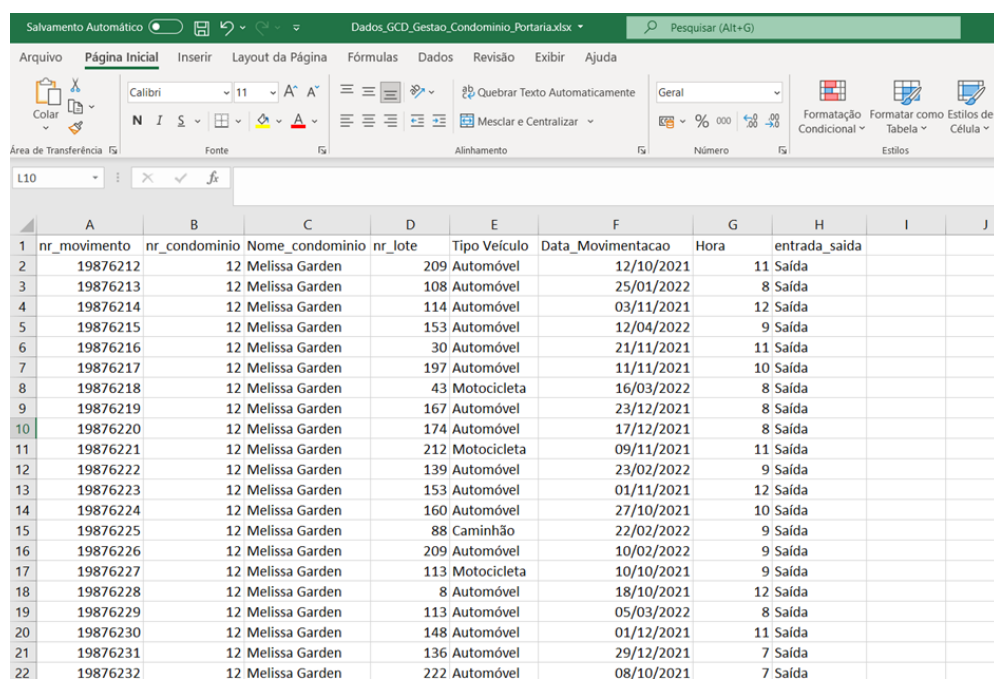
HIDG

Hugo, Leny, Vitor, Wallace

1.3 Desafio 2: gerando oportunidades com o uso dos dados

A relação entre Fred e os gestores dos condomínios que pretendem utilizar a solução de gestão de portaria é muito próxima. Muitos dos proprietários dos lotes em condomínio são empreendedores dos mais variados tipos, pessoas públicas, donos de pequenas, médias e grandes empresas, pessoas físicas que normalmente adquiriram seu lote ou mesmo sua residência.

Essa diversidade de tipos de pessoas utiliza diariamente as portarias do condomínio e a cada passagem um registro com a informação é confirmado. Veja um pequeno exemplo de layout e com dados gerados



	A	B	C	D	E	F	G	H	I	J
1	nr_movimento	nr_condominio	Nome_condominio	nr_lote	Tipo Veículo	Data_Movimentacao	Hora	entrada_saida		
2	19876212	12	Melissa Garden	209	Automóvel	12/10/2021	11	Saída		
3	19876213	12	Melissa Garden	108	Automóvel	25/01/2022	8	Saída		
4	19876214	12	Melissa Garden	114	Automóvel	03/11/2021	12	Saída		
5	19876215	12	Melissa Garden	153	Automóvel	12/04/2022	9	Saída		
6	19876216	12	Melissa Garden	30	Automóvel	21/11/2021	11	Saída		
7	19876217	12	Melissa Garden	197	Automóvel	11/11/2021	10	Saída		
8	19876218	12	Melissa Garden	43	Motocicleta	16/03/2022	8	Saída		
9	19876219	12	Melissa Garden	167	Automóvel	23/12/2021	8	Saída		
10	19876220	12	Melissa Garden	174	Automóvel	17/12/2021	8	Saída		
11	19876221	12	Melissa Garden	212	Motocicleta	09/11/2021	11	Saída		
12	19876222	12	Melissa Garden	139	Automóvel	23/02/2022	9	Saída		
13	19876223	12	Melissa Garden	153	Automóvel	01/11/2021	12	Saída		
14	19876224	12	Melissa Garden	160	Automóvel	27/10/2021	10	Saída		
15	19876225	12	Melissa Garden	88	Caminhão	22/02/2022	9	Saída		
16	19876226	12	Melissa Garden	209	Automóvel	10/02/2022	9	Saída		
17	19876227	12	Melissa Garden	113	Motocicleta	10/10/2021	9	Saída		
18	19876228	12	Melissa Garden	8	Automóvel	18/10/2021	12	Saída		
19	19876229	12	Melissa Garden	113	Automóvel	05/03/2022	8	Saída		
20	19876230	12	Melissa Garden	148	Automóvel	01/12/2021	11	Saída		
21	19876231	12	Melissa Garden	136	Automóvel	29/12/2021	7	Saída		
22	19876232	12	Melissa Garden	222	Automóvel	08/10/2021	7	Saída		

O seu desafio será criar uma história por meio de dados utilizando a narrativa de dados (storytelling). Descreva o que foi utilizado para deixar os dados adequados para os usuários finais (visual, contexto para comunicação, narrativa, fontes de dados).

HIDG

Hugo, Leny, Vitor, Wallace

Sumário

Contextualizando o problema	3
Proposta StoryTelling:.....	4
Importando bibliotecas e lendo o arquivo de dados csv.....	5
Análise exploratória dos dados	6
Verificando valores missing.....	6
Verificando os diferentes valores de cada uma das colunas.....	7
Adicionando nova coluna contendo o dia da semana, indo de domingo sendo 0 e 6 sendo sábado.....	7
Verificação da quantidade de dados com base no ano	7
Verificação dos dados com base nos meses	8
Retirando dados de mês indesejado.....	8
Quantidade de entradas e saídas	8
Contagem dos horários de entrada e saída	8
Quantidade de valores dos 2 condomínios	9
Quantidade de lotes	9
Retirada de dados incorretos.....	9
Frequências e histogramas.....	10
Outliers.....	12
Análise de séries temporais com base nos últimos meses	13
Moradia específica.....	18
Conclusão.....	23

HIDG

Hugo, Leny, Vitor, Wallace

Contextualizando o problema

Temos o arquivo relacionado a portaria de 2 diferentes condomínios: Melissa Garden e Flowers Queen, no caso a primeira opção com muito mais valores que a segunda, então consideraremos inicialmente análises aplicadas em conjunto por esses condomínios compartilharem as mesmas colunas e possuir poucos valores no segundo condomínio. Inicialmente podemos notar a distribuição dos valores que temos em diferentes colunas, dentre as quais com valores quantitativos relevantes para medir o nível de segurança dos moradores de determinado lote (no caso a coluna hora e dia da semana); podemos utilizar dessas colunas para verificar os padrões de entrada e saída dos moradores, para quando ocorrer uma divergência desses dados, a portaria seja sinalizada. A partir desse pressuposto podemos procurar por padrões nos dados disponibilizados, e nisso conseguiremos chegar a certas conclusões, dentre as quais listadas abaixo, que serão demonstradas pelos códigos mais adiante do relatório:

- Todos os lotes, sem exceção, não passam de determinado horário para saída (12h) nem para entrada do condomínio (21h).
- O nível de movimentação pela portaria varia com base nos dias da semana, sendo domingo o dia com menos movimentação no total, variando entre os dias da semana e tendo sexta-feira com mais movimento total.
- O mês com menos movimentação foi de fevereiro de 2022 (3568) enquanto o com mais movimentação foi janeiro de 2022 (3808), e com pouca disparidade entre os demais meses.
- A distribuição desses horários é similar, podendo variar apenas de lote para lote.
- No mês de fevereiro de 2022 teve uma grande queda no número de entradas e saídas na portaria do Melissa Garden, o que impulsiona a acontecer o mesmo em relação aos tipos de veículos utilizados nos últimos meses, por exemplo.
- No Flowers Queen essa frequência se manteve consistente nos últimos meses.



HIDG

Hugo, Leny, Vitor, Wallace

Proposta StoryTelling:

Checaremos a segurança dos moradores dos 2 condomínios por meio de algumas métricas, dentre as quais:

- Outliers nos horários de entrada e saída dos moradores
- Os dias da semana que as pessoas mais costumam sair e possíveis inconsistências
- Os tipos de veículos mais utilizados pelos moradores
- Verificar a movimentação de ambos os condomínios nos últimos meses
- Verificar possível correlação entre o dia da semana e o horário de entrada e saída

E para isso precisaremos seguir os seguintes passos para o processo de análise exploratória:

- Leitura do Dataframe
- Verificação dos dados de cada uma das colunas
- Contagem dos diferentes valores de cada coluna
- Retirada de possíveis inconsistências
- Adicionando nova coluna contendo o dia da semana
- Verificando histograma com os horários de entrada e saída dos moradores
- Verificação de outliers
- Gráfico contendo a frequência de entrada e saída dos moradores nos últimos meses
- Gráfico contendo a frequência do uso de diferentes automóveis
- Realização de análises estatísticas de uma única moradia.
- Verificando possível correlação entre o horário de entrada e saída e o dia da semana

HIDG

Hugo, Leny, Vitor, Wallace

Importando bibliotecas e lendo o arquivo de dados csv

```
0s import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import datetime

[2] from statsmodels.tsa.stattools import adfuller
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.arima_model import ARIMA
from sklearn.metrics import mean_squared_error, mean_absolute_percentage_error, mean_absolute_error

/usr/local/lib/python3.7/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning: pandas.util.testing
import pandas.util.testing as tm

[3] df_portaria = pd.read_excel('Dados_GCD_Gestao_Condominio_Portaria.xlsx')
```

FIAP

HIDG

Hugo, Leny, Vitor, Wallace

Análise exploratória dos dados

```

[4] df_portaria.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 26142 entries, 0 to 26141
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   nr_movimento           26142 non-null  int64  
1   nr_condominio          26142 non-null  int64  
2   Nome_condominio        26142 non-null  object  
3   nr_lote                26142 non-null  int64  
4   Tipo Veículo           26142 non-null  object  
5   Data_Movimentacao      26142 non-null  datetime64[ns]
6   Hora                   26142 non-null  int64  
7   entrada_saida          26142 non-null  object  
dtypes: datetime64[ns](1), int64(4), object(3)
memory usage: 1.6+ MB

```

Temos todos os tipos de valores adequadamente organizados em cada coluna, com valores de texto, número inteiro e campo de data.

Verificando valores missing

```

[5] for col in df_portaria.columns:
    pct_missing = np.mean(df_portaria[col].isnull())
    print('{} - {}'.format(col, round(pct_missing*100)))

```

```

nr_movimento - 0%
nr_condominio - 0%
Nome_condominio - 0%
nr_lote - 0%
Tipo Veículo - 0%
Data_Movimentacao - 0%
Hora - 0%
entrada_saida - 0%

```

```

[6] df_portaria.isnull().sum()

```

```

nr_movimento      0
nr_condominio      0
Nome_condominio    0
nr_lote            0
Tipo Veículo       0
Data_Movimentacao  0
Hora               0
entrada_saida      0
dtype: int64

```

Não temos nenhum valor missing em nenhuma das colunas, então não é necessário nenhum tipo de tratamento nesse sentido.

HIDG

Hugo, Leny, Vitor, Wallace

Verificando os diferentes valores de cada uma das colunas

```
[7] for col in df_portaria.columns:
    print('{} : {}'.format(col, df_portaria[col].unique()))

nr_movimento : [19876212 19876213 19876214 ... 6681026 6681027 6681028]
nr_condominio : [12 17]
Nome_condominio : ['Melissa Garden' 'Flowers Queen']
nr_lote : [209 108 114 153 30 197 43 167 174 212 139 160 88 113 8 148 136 222
142 40 215 81 65 6 98 166 159 34 130 162 12 33 107 32 27 207
203 165 94 64 71 195 132 20 83 45 55 74 127 2 183 196 179 134
56 154 147 156 1 152 75 199 90 204 66 4 70 129 73 125 26 17
82 28 124 29 201 180 53 77 175 46 173 170 76 163 126 59 61 193
103 44 182 91 50 52 141 15 138 63 206 42 191 145 22 205 68 41
111 47 172 194 37 133 99 67 14 25 128 208 93 135 176 189 62 116
106 36 188 202 9 23 35 5 109 57 89 11 84 31 192 7 123 96
157 217 101 117 18 95 177 104 210 86 13 39 187 115 151 198 49 102
143 80 200 131 110 213 69 60 118 48 58 87 150 214 216 3 186 211
161 144 137 218 155 181 164 92 119 221 185 158 16 38 149 21 97 171
178 219 10 122 146 19 78 168 220 54 190 112 169 120 24 121 51 85
100 105 79 184 140 72]
Tipo Veículo : ['Automóvel' 'Motocicleta' 'Caminhão' 'v']
Data_Movimentacao : ['2021-10-12T00:00:00.000000000' '2022-01-25T00:00:00.000000000'
'2021-11-03T00:00:00.000000000' '2022-04-12T00:00:00.000000000']
Hora : [11 8 12 9 10 7 21 18 19 20 17]
entrada_saida : ['Saída' 'Entrada']
```

Podemos notar que temos o número do condomínio de acordo com os dois nomes do condomínio: Melissa Garden (12) e Flowers Queen (17).

```
[9] df_portaria.drop_duplicates()
```

Também não foi encontrado nenhum valor duplicado, então não há necessidade de tratamento nesse sentido.

Adicionando nova coluna contendo o dia da semana, indo de domingo sendo 0 e 6 sendo sábado

```
[11] df_portaria['dia_semana'] = df_portaria['Data_Movimentacao'].dt.dayofweek
```

```
[12] df_portaria
```

	nr_movimento	nr_condominio	Nome_condominio	nr_lote	Tipo Veículo	Data_Movimentacao	Hora	entrada_saida	dia_semana
0	19876212	12	Melissa Garden	209	Automóvel	2021-10-12	11	Saída	1
1	19876213	12	Melissa Garden	108	Automóvel	2022-01-25	8	Saída	1
2	19876214	12	Melissa Garden	114	Automóvel	2021-11-03	12	Saída	2
3	19876215	12	Melissa Garden	153	Automóvel	2022-04-12	9	Saída	1
4	19876216	12	Melissa Garden	30	Automóvel	2021-11-21	11	Saída	6
...

Verificação da quantidade de dados com base no ano

```
[15] df_portaria['Data_Movimentacao'].groupby(df_portaria.Data_Movimentacao.dt.to_period("Y")).agg('count')

Data_Movimentacao
2021    11254
2022    14888
Freq: A-DEC, Name: Data_Movimentacao, dtype: int64
```

HIDG

Hugo, Leny, Vitor, Wallace

Verificação dos dados com base nos meses

```

✓ [17] df_portaria['Data_Movimentacao'].groupby(df_portaria.Data_Movimentacao.dt.to_period("M")).agg('count')
0a
Data_Movimentacao
2021-10      3800
2021-11      3708
2021-12      3746
2022-01      3808
2022-02      3538
2022-03      3654
2022-04      3768
2022-05       120
Freq: M, Name: Data_Movimentacao, dtype: int64

```

Retirando dados de mês indesejado

Necessária a retirada dos dados do mês de maio, por apresentar poucos dados para a análise de séries temporais, impactando nos resultados, e como os dados foram disponibilizados no início do mês, é compatível com a base de dados apresentada, então consideraremos apenas os dados do mês de abril e anteriores.

```

✓ [18] df_portaria = df_portaria.loc[df_portaria.Data_Movimentacao.dt.to_period("M") != '2022-05']
0a

```

Quantidade de entradas e saídas

Vemos que ainda assim possuímos a mesma quantidade de valores tanto para entrada quanto para saída

```

✓ [19] df_portaria['entrada_saida'].value_counts()
0a
Saída      13011
Entrada    13011
Name: entrada_saida, dtype: int64

```

Contagem dos horários de entrada e saída

Verificando os principais horários de entrada e saída do condomínio, e assim percebemos que tem uma maior movimentação no período das 17 até as 20, porém do período das 7 às 12 tem uma hora a mais que o período vespertino-noturno, indicando uma movimentação parecida em ambos os períodos.

```

✓ [110] df_portaria['Hora'].value_counts()
0a
19      2619
20      2618
18      2612
21      2608
17      2548
9        2243
11       2178
8        2178
10       2160
12       2145
7        2105
Name: Hora, dtype: int64

```

```

✓ [111] df_portaria['Hora'].value_counts().plot(kind='bar')
0a
plt.title('Horários de pico das saídas e entrada das residências')
plt.show()

```



HIDG

Hugo, Leny, Vitor, Wallace

Quantidade de valores dos 2 condomínios

Com os outros valores percebemos valores de duas portarias diferentes, uma delas do Melissa Garden e outra do Flowers Garden, com quase um quarto da quantidade de linhas do primeiro condomínio.

```
✓ [21] df_portaria['nr_condominio'].value_counts()
```

```
12    19988
17     6034
Name: nr_condominio, dtype: int64
```

```
✓ [22] df_portaria['Nome_condominio'].value_counts()
```

```
Melissa Garden    19988
Flowers Queen     6034
Name: Nome_condominio, dtype: int64
```

Quantidade de lotes

E nisso percebemos 222 diferentes lotes, variando na quantidade de entradas e saídas de cada um, indo de 58 entradas e saídas até 304 no mais movimentado.

```
✓ [24] df_portaria['nr_lote'].value_counts()
```

```
9      304
7      296
26     270
1      270
11     264
...
118     62
201     62
73      60
179     58
158     58
Name: nr_lote, Length: 222, dtype: int64
```

Retirada de dados incorretos

Foram encontradas inconsistências referentes ao tipo de veículo em alguns valores da coluna “Tipo Veículo”, que tinham o valor ‘v’, então como eram poucas linhas e de diferentes lotes, foi optado pela remoção desses dados para posteriores análises dessa coluna.

```
✓ [25] df_portaria.loc[df_portaria['Tipo Veículo'] == 'v']
```

	nr_movimento	nr_condominio	Nome_condominio	nr_lote	Tipo Veículo	Data_Movimentacao	Hora	entrada_saida	dia_semana
227	19876439	12	Melissa Garden	147	v	2022-04-07	9	Saída	3
955	19877167	12	Melissa Garden	69	v	2022-01-19	8	Saída	2
17171	19981762	12	Melissa Garden	159	v	2021-10-31	17	Entrada	6
17591	19982182	12	Melissa Garden	110	v	2022-04-24	18	Entrada	6
19831	19984422	12	Melissa Garden	132	v	2022-04-06	19	Entrada	2
20643	19985234	12	Melissa Garden	104	v	2021-12-12	21	Entrada	6
22144	19986735	12	Melissa Garden	172	v	2022-01-05	20	Entrada	2
24188	6679075	17	Flowers Queen	28	v	2021-11-15	20	Entrada	0

```
✓ [27] df_portaria = df_portaria.loc[df_portaria['Tipo Veículo'] != 'v']
```

```
✓ [28] df_portaria.loc[df_portaria['Tipo Veículo'] == 'v']
```

```
nr_movimento nr_condominio Nome_condominio nr_lote Tipo Veículo Data_Movimentacao Hora entrada_saida dia_semana
```

HIDG

Hugo, Leny, Vitor, Wallace

Frequências e histogramas

Notamos também os dias da semana que ocorre maior movimentação, levando destaque para sexta-feira (5) com relativo aumento de movimentação em comparação ao dia menos movimentado, que foi domingo (0), com uma diferença de 163 entradas e 162 saídas em relação à domingo:

```
[5] df_portaria.groupby(['dia_semana', 'entrada_saida'])['dia_semana'].count()
```

```
dia_semana  entrada_saida
0          Entrada      1777
           Saída      1778
1          Entrada      1797
           Saída      1797
2          Entrada      1926
           Saída      1927
3          Entrada      1854
           Saída      1853
4          Entrada      1855
           Saída      1855
5          Entrada      1940
           Saída      1940
6          Entrada      1856
           Saída      1859
Name: dia_semana, dtype: int64
```

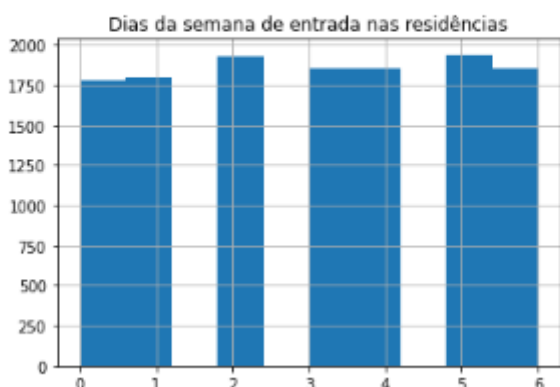
Verificando padrões relacionados aos horários de entrada e saída dos moradores em 2, um para entrada e outro para saída.

```
[114] dados_entrada = df_portaria.loc[df_portaria['entrada_saida'] == 'Entrada']
```

```
[115] dados_entrada['Hora'].hist()
plt.title("Horários de entrada nas residências")
plt.show()
```



```
[116] dados_entrada['dia_semana'].hist()
plt.title("Dias da semana de entrada nas residências")
plt.show()
```



Notamos um padrão muito semelhante em diferentes horários de entrada e nos dias da semana, com algumas variações nos dias,

HIDG

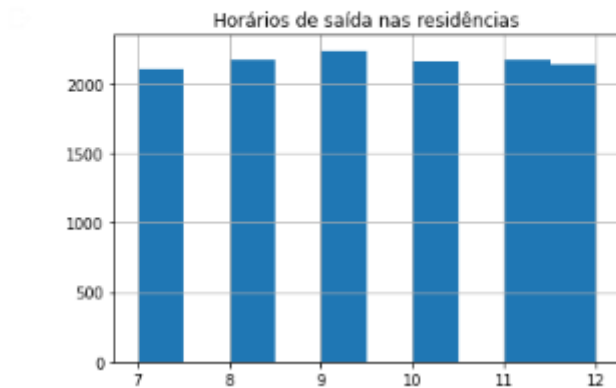
Hugo, Leny, Vitor, Wallace

como foi citado anteriormente, e encontramos o mesmo padrão nos dados de saída dos lotes:

Análises referentes às saídas registradas na portaria

```
[117] dados_saida = df_portaria.loc[df_portaria['entrada_saida'] == 'Saída']
```

```
[118] dados_saida['Hora'].hist()
plt.title("Horários de saída nas residências")
plt.show()
```



```
[119] dados_saida['dia_semana'].hist()
plt.title("Dias da semana de saída nas residências, (0: Domingo, 6:Sábado)")
plt.show()
```



HIDG

Hugo, Leny, Vitor, Wallace

Outliers

Após a verificação de outliers notamos que não há nenhum valor discrepante na coluna “Hora”, o que indica que esses moradores possuem horários fixos de entrada e saída, assim sendo mais simples de prever um comportamento anormal vindo desses dados, com alguns exemplos que serão demonstrados abaixo.

```
✓ dados_entrada.shape
```

```
(13005, 9)
```

```
✓ [66] df_unico = dados_entrada
```

```
✓ [67] df_unico['std*1.96'] = 1.96*dados_entrada['Hora'].std()
df_unico['Outlier'] = abs(dados_entrada['Hora'] - dados_entrada['Hora'].mean()) > df_unico['std*1.96']
outliers_dados_entrada = dados_entrada.loc[dados_entrada['Outlier'] == True]
df_unico['Outlier'].value_counts()[False]
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
13005
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
This is separate from the ipykernel package so we can avoid doing imports until

```
13009
```



HIDG

Hugo, Leny, Vitor, Wallace

Análise de séries temporais com base nos últimos meses

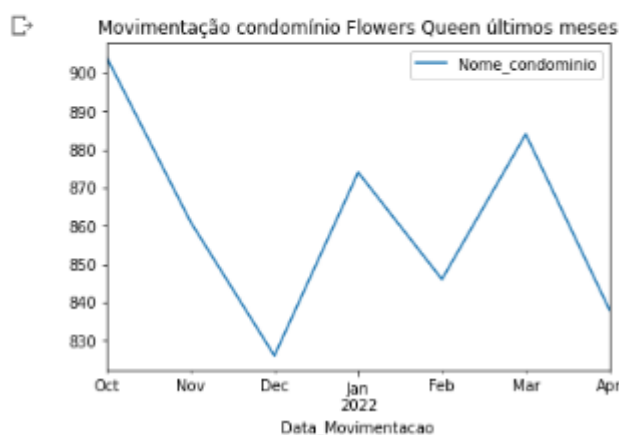
Começando com a movimentação do condomínio Melissa Garden nos últimos meses, onde notamos uma queda brusca desses valores do mês de janeiro para fevereiro, e estabilizando posteriormente até o mês de abril, o que impacta também na leitura dos próximos gráficos que são influenciados por esses valores.

```
[122] melissa_garden = df_portaria.loc[df_portaria['Nome_condominio'] == 'Melissa Garden']
      melissa_garden_analises = melissa_garden[['Data_Movimentacao', 'Nome_condominio']]
      melissa_garden_analises.set_index('Data_Movimentacao', inplace = True)
      dados_melissa_garden_agg = melissa_garden_analises.groupby(pd.Grouper(freq='M')).count()
      dados_melissa_garden_agg.plot()
      plt.title('Movimentação últimos meses condomínio Melissa Garden')
      plt.show()
```



Enquanto no Flowers Queen temos essa queda no mês de fevereiro muito menor, porém tendo uma queda de maior relevância do mês de outubro até dezembro de 2021.

```
flowers_queen = df_portaria.loc[df_portaria['Nome_condominio'] == 'Flowers Queen']
flowers_queen_analises = flowers_queen[['Data_Movimentacao', 'Nome_condominio']]
flowers_queen_analises.set_index('Data_Movimentacao', inplace = True)
dados_flowers_queen_agg = flowers_queen_analises.groupby(pd.Grouper(freq='M')).count()
dados_flowers_queen_agg.plot()
plt.title('Movimentação condomínio Flowers Queen últimos meses')
plt.show()
```



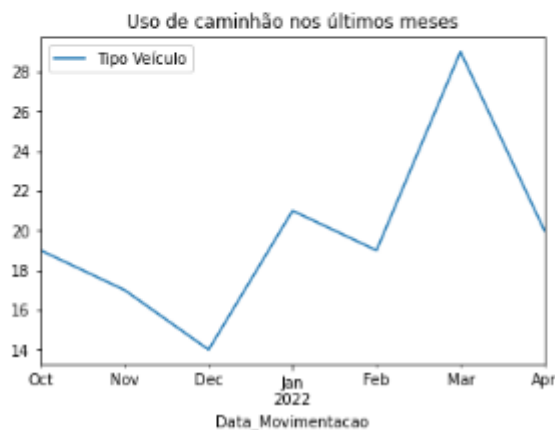
Com o uso de caminhão em ambos os condomínios, notamos poucos valores e que houve uma crescente para o mês de março que se estabilizou no mês de abril.

HIDG

Hugo, Leny, Vitor, Wallace

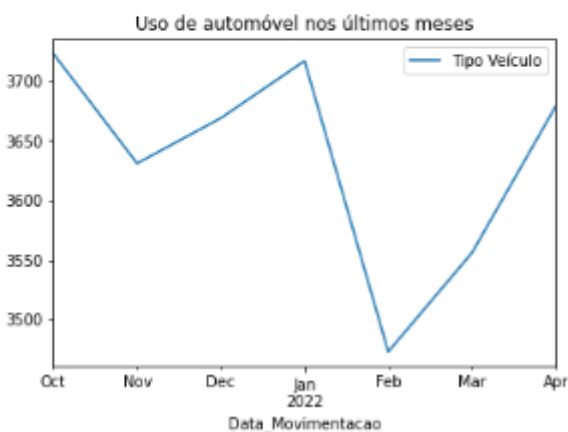
```
✓ [120] caminhao = df_portaria.loc[df_portaria['Tipo Veículo'] == 'Caminhão']
```

```
✓ [123] caminhao_analises = caminhao[['Data_Movimentacao', 'Tipo Veículo']]
caminhao_analises.set_index('Data_Movimentacao', inplace = True)
dados_caminhao_agg = caminhao_analises.groupby(pd.Grouper(freq='M')).count()
dados_caminhao_agg.plot()
plt.title('Uso de caminhão nos últimos meses')
plt.show()
```



Com o uso de automóvel notamos grande semelhança com o gráfico de movimentação do condomínio Melissa Garden, por se tratar da grande maioria que utiliza automóveis em relação aos outros meios de transporte encontrados.

```
✓ [121] automovel = df_portaria.loc[df_portaria['Tipo Veículo'] == 'Automóvel']
automovel_analises = automovel[['Data_Movimentacao', 'Tipo Veículo']]
automovel_analises.set_index('Data_Movimentacao', inplace = True)
dados_automovel_agg = automovel_analises.groupby(pd.Grouper(freq='M')).count()
dados_automovel_agg.plot()
plt.title('Uso de automóvel nos últimos meses')
plt.show()
```



Também temos esses valores impactando no uso de motocicletas pelos condomínios, como é possível notar no gráfico abaixo:

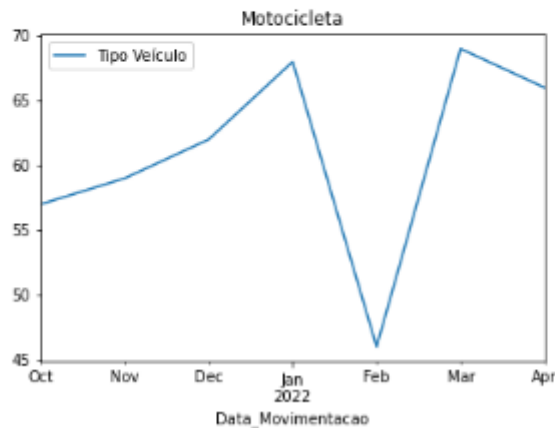
HIDG

Hugo, Leny, Vitor, Wallace

```

moto = df_portaria.loc[df_portaria['Tipo Veículo'] == 'Motocicleta']
moto_analises = moto[['Data_Movimentacao', 'Tipo Veículo']]
moto_analises.set_index('Data_Movimentacao', inplace = True)
dados_moto_agg = moto_analises.groupby(pd.Grouper(freq='M')).count()
dados_moto_agg.plot()
plt.title('Motocicleta')
plt.show()

```

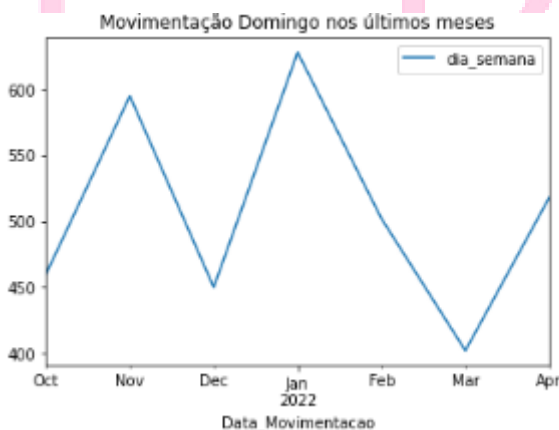


Também sendo possível observar essa movimentação se baseando nos dias da semana, como podemos observar a seguir:

```

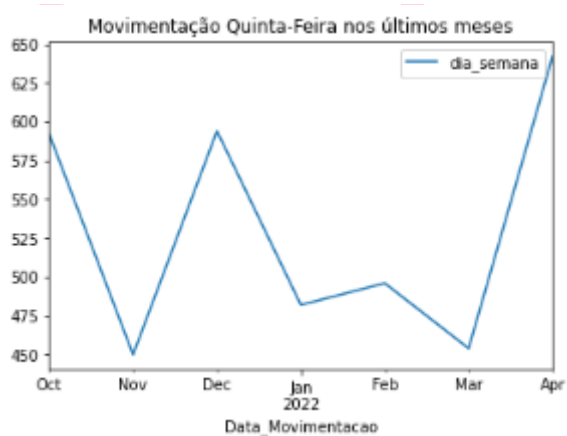
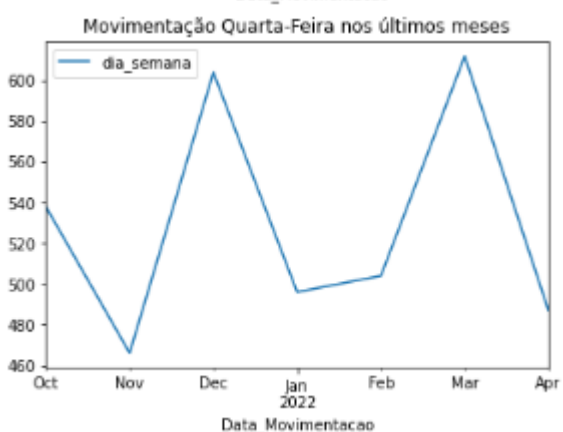
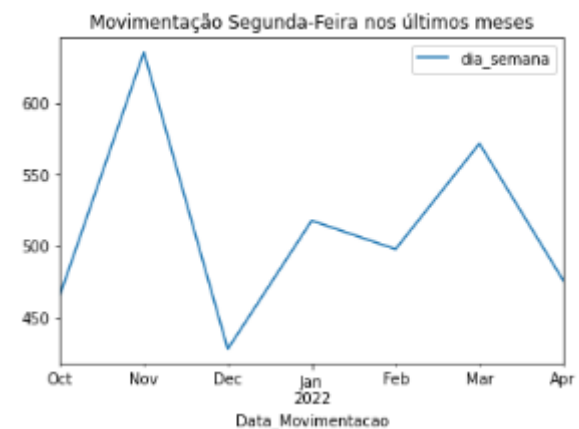
[103] dias_semana = ['Domingo', 'Segunda-Feira', 'Terça-Feira', 'Quarta-Feira', 'Quinta-Feira', 'Sexta-Feira', 'Sábado']
for i in range(len(dias_semana)):
    flowers_queen = df_portaria.loc[df_portaria['dia_semana'] == i]
    flowers_queen_analises = flowers_queen[['Data_Movimentacao', 'dia_semana']]
    flowers_queen_analises.set_index('Data_Movimentacao', inplace = True)
    dados_flowers_queen_agg = flowers_queen_analises.groupby(pd.Grouper(freq='M')).count()
    dados_flowers_queen_agg.plot()
    plt.title('Movimentação {} nos últimos meses'.format(dias_semana[i]))
    plt.show()

```



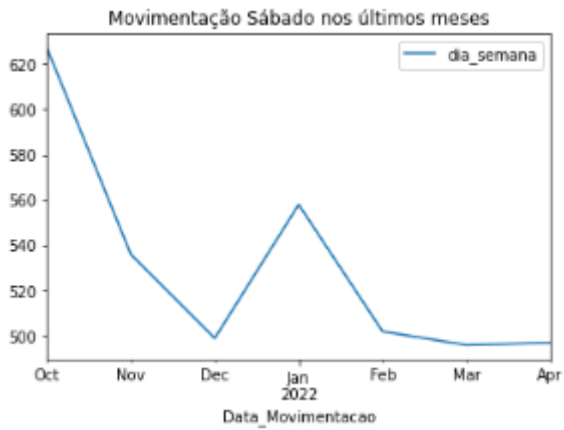
HIDG

Hugo, Leny, Vitor, Wallace



HIDG

Hugo, Leny, Vitor, Wallace



Sendo possível interpretar demasiada divergência na frequência com que as pessoas saem durante a semana, em alguns meses tendo um pico maior em alguns dias e menor em outros, explicitando melhor o comportamento dos moradores desses condomínios.

FIAP

HIDG

Hugo, Leny, Vitor, Wallace

Moradia específica

Analisaremos os dados de uma moradia escolhida aleatoriamente para demonstrar a detecção de possíveis inconsistências nos horários de entrada e saída com base nos dias da semana desse determinado lote, o que poderia ajudar na segurança caso fosse encontrado alguma anomalia nesses dados, podendo ser repassado essa informação para a portaria entrar em alerta e tomar possíveis medidas que garantam a segurança do condomínio.

Lote escolhido: 209

```
[46] casa1 = df_portaria.loc[df_portaria['nr_lote'] == 209]
      casa1
```

	nr_movimento	nr_condominio	Nome_condominio	nr_lote	Tipo Veículo	Data_Movimentacao	Hora	entrada_saida	dia_semana
0	19876212	12	Melissa Garden	209	Automóvel	2021-10-12	11	Saída	1
14	19876226	12	Melissa Garden	209	Automóvel	2022-02-10	9	Saída	3
82	19876294	12	Melissa Garden	209	Automóvel	2022-04-04	11	Saída	0
236	19876448	12	Melissa Garden	209	Automóvel	2021-10-05	9	Saída	1
250	19876462	12	Melissa Garden	209	Automóvel	2022-03-15	9	Saída	1
...
22422	19987013	12	Melissa Garden	209	Automóvel	2022-01-10	19	Entrada	0
22582	19987173	12	Melissa Garden	209	Automóvel	2021-11-05	21	Entrada	4
22735	19987326	12	Melissa Garden	209	Automóvel	2021-10-30	20	Entrada	5
23006	19987597	12	Melissa Garden	209	Automóvel	2021-11-13	19	Entrada	5
23076	19987667	12	Melissa Garden	209	Automóvel	2022-02-28	19	Entrada	0

88 rows x 9 columns

Separando os valores de entrada e saída para facilitar as análises

```
[50] entrada_casa1 = casa1.loc[casa1['entrada_saida'] == 'Entrada']
      saida_casa1 = casa1.loc[casa1['entrada_saida'] == 'Saída']
```

Notamos que a(s) pessoa(s) desse determinado lote tem maior frequência de saídas entre as 7h e 9h, e menos das 10h até 12h.

```
[51] saida_casa1['Hora'].hist()
      plt.title("Principais horários de saída do lote 209")
      plt.show()
```



HIDG

Hugo, Leny, Vitor, Wallace

Enquanto de entrada encontramos certa estabilidade nos diferentes horários, com 20h sendo o principal horário de entrada e 18 o de menos frequência.

```
✓ entrada_casa1['Hora'].hist()
0s plt.title("Principais horários de entrada do lote 209")
plt.show()
```



Notamos também que nesse determinado lote existe maior frequência de entrada e saída na sexta-feira (5), e de menor movimentação no sábado (6)

```
✓ [130] casa1['dia_semana'].hist()
0s plt.title("Principais dias da semana de entrada e saída do lote 209")
plt.show()
```



Podemos verificar também se existe uma possível correlação entre o dia da semana e o horário de entrada e saída nesse determinado lote:

```
[ ] entrada_casa1['Hora'].corr(entrada_casa1['dia_semana'])
```

0.018503159047701244

Temos um valor considerado bem baixo (1,8%), então podemos descartar essa possibilidade nessa moradia.

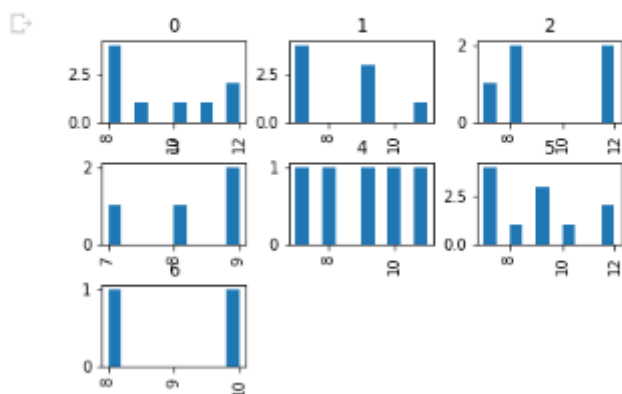
HIDG

Hugo, Leny, Vitor, Wallace

Também podemos tirar algumas métricas relacionadas aos horários de entrada e saída com base nos dias da semana, no caso temos os seguintes histogramas, indicando a frequência desses horários com base no dia da semana:

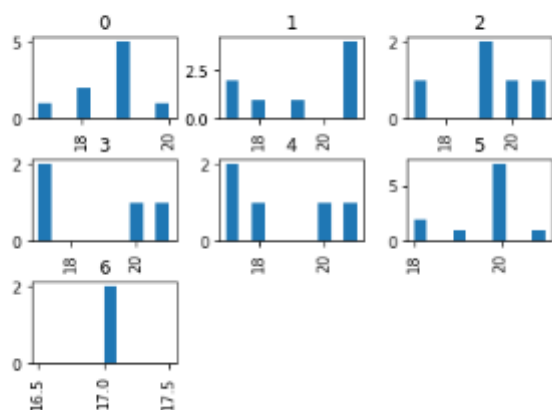
Dados de saída

```
saida_casa1.hist(by='dia_semana', column='Hora')
plt.show()
```



Dados de entrada

```
[ ] entrada_casa1.hist(by='dia_semana', column='Hora')
plt.show()
```

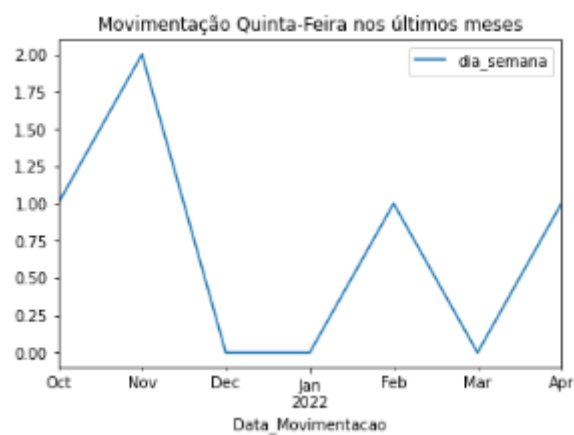
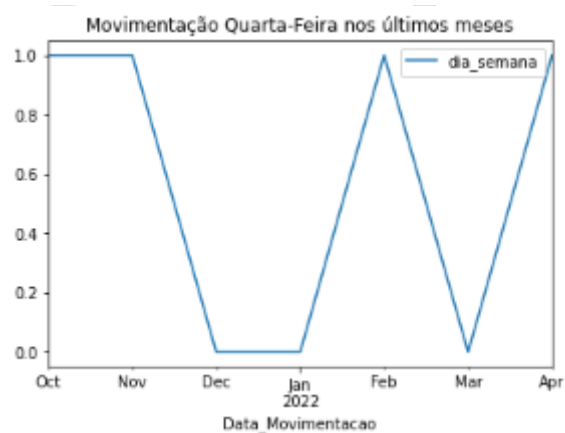
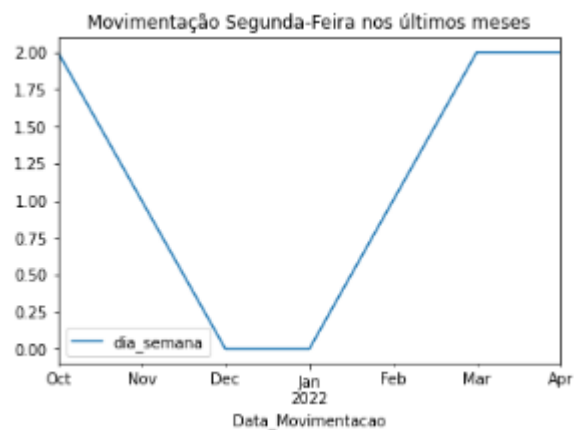


E, por fim, podemos realizar a mesma análise de séries temporais anteriormente para entender melhor o comportamento desse lote com base nos últimos meses e em diferentes dias da semana do lote 209:

```
[116] dias_semana = ['Domingo', 'Segunda-Feira', 'Terça-Feira', 'Quarta-Feira', 'Quinta-Feira', 'Sexta-Feira', 'Sábado']
for i in range(len(dias_semana)):
    flowers_queen = saida_casa1.loc[saida_casa1['dia_semana'] == i]
    flowers_queen_analises = flowers_queen[['Data_Movimentacao', 'dia_semana']]
    flowers_queen_analises.set_index('Data_Movimentacao', inplace = True)
    dados_flowers_queen_agg = flowers_queen_analises.groupby(pd.Grouper(freq='M')).count()
    dados_flowers_queen_agg.plot()
    plt.title('Movimentação {} nos últimos meses'.format(dias_semana[i]))
    plt.show()
```

HIDG

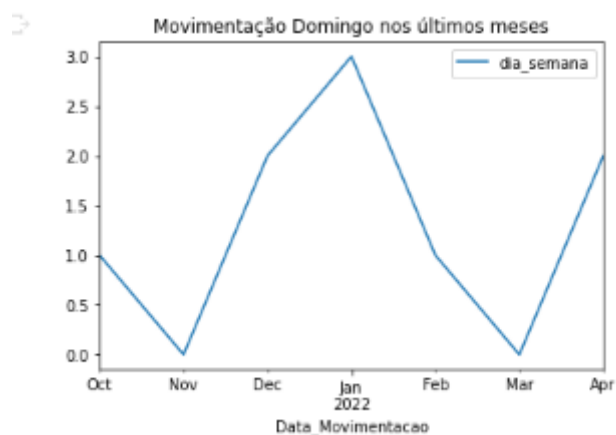
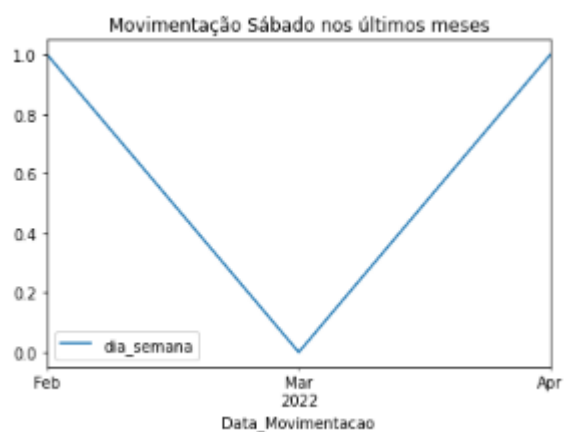
Hugo, Leny, Vitor, Wallace



FIAP

HIDG

Hugo, Leny, Vitor, Wallace



FIAP

HIDG

Hugo, Leny, Vitor, Wallace

Conclusão

Dessa maneira, obtivemos valores que podem contribuir para a tomada de decisão na administração desse condomínio, com métricas e gráficos relacionados aos padrões dos moradores e também com base nos últimos meses, de modo que caso seja encontrado determinado outlier em algum dos valores possa ser identificado o mais rápido possível e compreender a razão de tal acontecimento; assim, melhorando a segurança desse condomínio, e, caso necessário, compreendendo os padrões de determinado lote para o aumento dessa segurança sob medida.

Caso necessário também é possível a obtenção desses dados para a tentativa de previsão futura da **quantidade de acessos** que esse condomínio tem **semanalmente**, como pode ser exemplificado a seguir utilizando o modelo estatístico ARIMA em 7 semanas como teste e tendo também a porcentagem de acerto do modelo:

	Nome_condominio	previsoes	Pct_acerto
Data_Movimentacao			
2022-03-21	320	281.000000	87.812500
2022-03-28	312	219.779156	70.442037
2022-04-04	299	99.255681	33.195880
2022-04-11	335	173.831544	51.890013
2022-04-18	382	203.212816	53.197072
2022-04-25	318	174.191455	54.777187
2022-05-02	261	191.468796	73.359692

Assim, de tal forma podemos adaptar o uso do cliente em relação à forma como ele pretende utilizar esses dados a seu favor para melhor tomada de decisão de seu *Business*, que nesse caso tivemos como proposta de storytelling a suposição de inconsistência nos valores de entrada e saída de determinado lote, e entendendo a fundo o que possivelmente poderia ocasionar esse tipo de ocasião.