

▼ Carregar dados

```
pip install scipy==1.7.1
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/
Collecting scipy==1.7.1
  Downloading scipy-1.7.1-cp37-cp37m-manylinux_2_5_x86_64.manylinux1_x86_64.whl (28.5MB)
    |████████| 28.5 MB 1.5 MB/s
Requirement already satisfied: numpy<1.23.0,>=1.16.5 in /usr/local/lib/python3.7/dist-packages
Installing collected packages: scipy
  Attempting uninstall: scipy
    Found existing installation: scipy 1.4.1
    Uninstalling scipy-1.4.1:
      Successfully uninstalled scipy-1.4.1
ERROR: pip's dependency resolver does not currently take into account all the packages in the dependencies of the package being installed.
albumentations 0.1.12 requires imgaug<0.2.7,>=0.2.5, but you have imgaug 0.2.9 which is incompatible because:
  - imgaug 0.2.9 is incompatible with albumentations 0.1.12 (requirement conflict).
Successfully installed scipy-1.7.1
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
pip install pingouin
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/
Collecting pingouin
  Downloading pingouin-0.5.1.tar.gz (183 kB)
    |████████| 183 kB 4.9 MB/s
Requirement already satisfied: numpy>=1.19 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: scipy>=1.7 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: pandas>=1.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: matplotlib>=3.0.2 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: seaborn>=0.11 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: statsmodels>=0.13
  Downloading statsmodels-0.13.2-cp37-cp37m-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.1MB)
    |████████| 9.8 MB 14.3 MB/s
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: pandas_flavor>=0.2.0
  Downloading pandas_flavor-0.3.0-py3-none-any.whl (6.3 kB)
Requirement already satisfied: outdated
  Downloading outdated-0.2.1-py3-none-any.whl (7.5 kB)
Requirement already satisfied: tabulate in /usr/local/lib/python3.7/dist-packages (from pandas_flavor>=0.2.0)
Requirement already satisfied: pyparsing!=2.0.4,!>=2.1.2,!>=2.1.6,>=2.0.1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: xarray in /usr/local/lib/python3.7/dist-packages (from pandas_flavor>=0.2.0)
  Downloading pandas_flavor-0.2.0-py2.py3-none-any.whl (6.6 kB)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from pandas_flavor>=0.2.0)
```

```

Requirement already satisfied: packaging>=21.3 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: patsy>=0.5.2 in /usr/local/lib/python3.7/dist-packages
Collecting littleutils
  Downloading littleutils-0.2.2.tar.gz (6.6 kB)
Requirement already satisfied: requests in /usr/local/lib/python3.7/dist-packages (from littleutils)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: joblib>=0.11 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (from littleutils)
Building wheels for collected packages: pingouin, littleutils
  Building wheel for pingouin (setup.py) ... done
    Created wheel for pingouin: filename=pingouin-0.5.1-py3-none-any.whl size=194002 stored in /root/.cache/pip/wheels/74/a7/b6/5b5f03b0395fdada1f45d3aed76e
  Stored in directory: /root/.cache/pip/wheels/74/a7/b6/5b5f03b0395fdada1f45d3aed76e
  Building wheel for littleutils (setup.py) ... done
    Created wheel for littleutils: filename=littleutils-0.2.2-py3-none-any.whl size=704 stored in /root/.cache/pip/wheels/d6/64/cd/32819b511a488e4993f2fab909a9f
  Stored in directory: /root/.cache/pip/wheels/d6/64/cd/32819b511a488e4993f2fab909a9f
Successfully built pingouin littleutils
Installing collected packages: littleutils, statsmodels, pandas-flavor, outdated, pingouin
  Attempting uninstall: statsmodels
    Found existing installation: statsmodels 0.10.2
    Uninstalling statsmodels-0.10.2:
      Successfully uninstalled statsmodels-0.10.2
Successfully installed littleutils-0.2.2 outdated-0.2.1 pandas-flavor-0.2.0 pingouin-0.5.1

```

```

"""
importando libs
scipy. a stats: associar dados
pingouin: associar dados com mais detalhes
pandas: manipulação de dados
numpy: manipulação de dados numericos
"""


```

```

"""
import seaborn as sns
import matplotlib.pyplot as plt
from scipy import stats as st
import pingouin as pg

import pandas as pd
import math
import numpy as np
from datetime import datetime

```

▼ Carregando arquivos

```

"""
importando dados da empresa
"""

dados_empresaStage1 = pd.read_csv('/content/drive/MyDrive/FIAP-ON/fase4/dw_gestao_condominio')

```

▼ Tratamento e transformação

`dados_empresaStage1.dtypes`

```
ID_FATO                      int64
ID_TEMPO                     int64
ID_DM_PROPRIETARIO           int64
ID_DM_MOVIMENTACAO          int64
ID_DM_CONDOMINIO             int64
ID_DM_LOTE                   int64
total_mvt_dia                int64
total_mvt_quinzena           int64
total_mvt_trimestre          int64
total_mvt_semestre            int64
total_mvt_ano                 int64
dt_movimentacao              datetime64[ns]
dia_semana                   int64
quinzena                      int64
trimestre                     int64
semestral                     int64
ano                           int64
ano_mes                        int64
NM_PROPRIETARIO               object
ST_GENERO_BIOLOGICO          object
tp_movimentacao               object
DS_RESUMIDA                   object
DS_LOGRADOURO_CONDOMINIO      object
NR_RESIDENCIA                  int64
DS_RAZAO_SOCIAL               object
NM_FANTASIA                   object
mes                            int64
ano_mes2                      datetime64[ns]
dtype: object
```

```
dados_empresaStage1['dt_movimentacao'] = pd.to_datetime(dados_empresaStage1["dt_movimentac
dados_empresaStage1['mes'] = dados_empresaStage1['dt_movimentacao'].dt.month
dados_empresaStage1['ano_mes2'] = pd.to_datetime(dados_empresaStage1["dt_movimentacao"]).dt
dados_empresaStage1.dtypes
```

```
ID_FATO                      int64
ID_TEMPO                     int64
ID_DM_PROPRIETARIO           int64
ID_DM_MOVIMENTACAO          int64
ID_DM_CONDOMINIO             int64
ID_DM_LOTE                   int64
total_mvt_dia                int64
total_mvt_quinzena           int64
total_mvt_trimestre          int64
total_mvt_semestre            int64
total_mvt_ano                 int64
dt_movimentacao              datetime64[ns]
dia_semana                   int64
quinzena                      int64
trimestre                     int64
semestral                     int64
```

```

ano                      int64
ano_mes                  int64
NM_PROPRIETARIO          object
ST_GENERO_BIOLOGICO     object
tp_movimentacao          object
DS_RESUMIDA               object
DS_LOGRADOURO_CONDOMINIO object
NR_RESIDENCIA             int64
DS_RAZAO_SOCIAL           object
NM_FANTASIA               object
mes                      int64
ano_mes2                 datetime64[ns]
dtype: object

```

▼ GERANDO GRAFICO

```

# total entrada e saida por condominio, condominio ano e mes
dados_empresaStageTotalPorCondominioAnoMes = dados_empresaStage1.groupby(['ano', 'mes']).a
dados_empresaStageTotalPorCondominioAno = dados_empresaStage1.groupby(['ano']).agg(total=p
dados_empresaStageTotalPorCondominioTotal = dados_empresaStage1.groupby(['NM_FANTASIA']).a

dados_empresaStageTotalPorCondominio = dados_empresaStage1.groupby(['NM_FANTASIA', 'ano_me

# total entrada por condominio, condominio ano e mes
dados_empresaStageTotalPorEntradaCondominio = dados_empresaStage1[dados_empresaStage1['tp_
dados_empresaStageTotalPorEntradaCondominioPorMes = dados_empresaStageTotalPorEntradaCondo
dados_empresaStageTotalPorEntradaCondominioAno = dados_empresaStageTotalPorEntradaCondomin
dados_empresaStageTotalPorEntradaCondominioTotal = dados_empresaStageTotalPorEntradaCondom

# total saida por condominio, condominio ano e mes
dados_empresaStageTotalPorSaidaCondominio = dados_empresaStage1[dados_empresaStage1['tp_mo
dados_empresaStageTotalPorSaidaCondominioPorMes = dados_empresaStageTotalPorSaidaCondomini
dados_empresaStageTotalPorSaidaCondominioAno = dados_empresaStageTotalPorSaidaCondominio.g
dados_empresaStageTotalPorSaidaCondominioTotal = dados_empresaStageTotalPorSaidaCondominio

# total veiculo na entrada e saida por condominio, condominio ano e mes
dados_empresaStageTotalPorVeiculoCondominioAnoMes = dados_empresaStage1.groupby(['DS_RESUM
dados_empresaStageTotalPorVeiculoCondominioTotal = dados_empresaStage1.groupby(['NM_FANTAS

dados_empresaStageTotalPorCondominioAnoMes['ano_mes'] = dados_empresaStageTotalPorCondominio
dados_empresaStageTotalPorEntradaCondominioPorMes['ano_mes'] = dados_empresaStageTotalPorEnt
dados_empresaStageTotalPorSaidaCondominioPorMes['ano_mes'] = dados_empresaStageTotalPorSai
dados_empresaStageTotalPorVeiculoCondominioAnoMes['ano_mes'] = dados_empresaStageTotalPorV

dados_empresaStageTotalPorCondominio

```

	NM_FANTASIA	ano_mes2	total
0	condonoo 1	2004-01-01	8
1	condonoo 1	2004-02-01	8
2	condonoo 1	2004-03-01	8
3	condonoo 1	2004-04-01	8
4	condonoo 1	2004-05-01	8
5	condonoo 1	2004-06-01	8
6	condonoo 1	2004-07-01	8
7	condonoo 1	2004-08-01	8
8	condonoo 1	2004-09-01	6
9	condonoo 1	2004-10-01	7
10	condonoo 1	2004-11-01	3
11	condonoo 2	2004-01-01	8
12	condonoo 2	2004-02-01	8
13	condonoo 2	2004-03-01	8
14	condonoo 2	2004-04-01	8
15	condonoo 2	2004-05-01	8
16	condonoo 2	2004-06-01	6
17	condonoo 2	2004-07-01	7
18	condonoo 2	2004-08-01	7
19	condonoo 2	2004-09-01	8
20	condonoo 2	2004-10-01	8
21	condonoo 2	2004-11-01	4
22	condonoo 3	2004-01-01	8
23	condonoo 3	2004-02-01	8
24	condonoo 3	2004-03-01	7
25	condonoo 3	2004-04-01	6
26	condonoo 3	2004-05-01	7
27	condonoo 3	2004-06-01	8
28	condonoo 3	2004-07-01	8
29	condonoo 3	2004-08-01	8
30	condonoo 3	2004-09-01	8
31	condonoo 3	2004-10-01	8
32	condonoo 3	2004-11-01	4

```

32 condonoo 3 2004-11-01 4
33 condonoo 4 2004-01-01 7
34 condonoo 4 2004-02-01 5
35 condonoo 4 2004-03-01 8
36 condonoo 4 2004-04-01 8
37 condonoo 4 2004-05-01 8
38 condonoo 4 2004-06-01 8
39 condonoo 4 2004-07-01 8
40 condonoo 4 2004-08-01 8
41 condonoo 4 2004-09-01 8

```

Gráfico por condomínio

```

43 conaonoo 4 2004-11-01 5

```

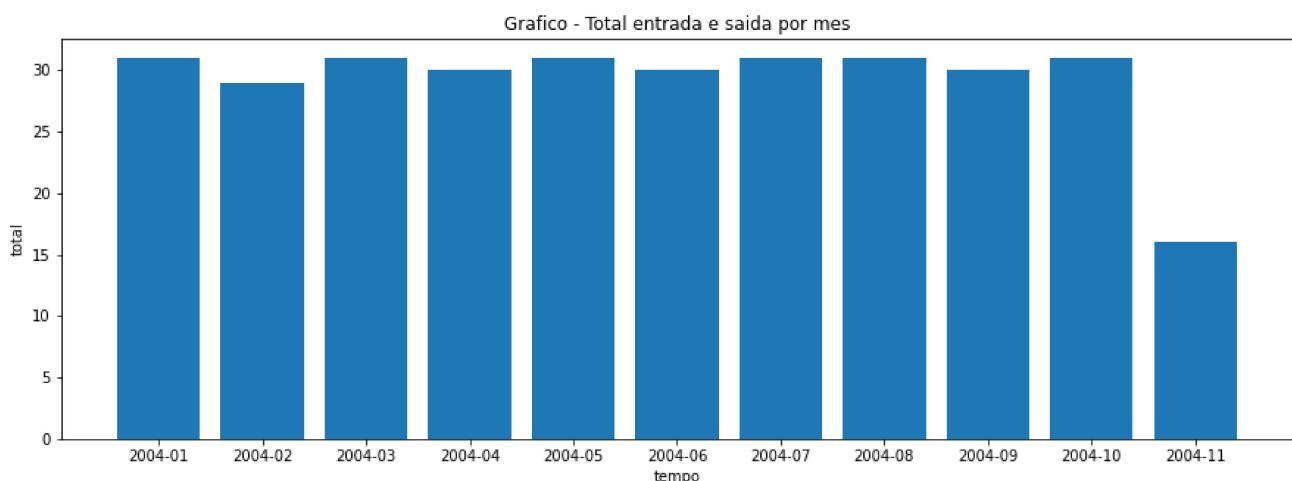
```

plt.rcParams["figure.figsize"] = (15,5)
plt.bar(dados_empresaStageTotalPorCondominioAnoMes[ 'ano_mes' ],dados_empresaStageTotalPorCo
plt.title('Grafico - Total entrada e saida por mes')
plt.xlabel('tempo')
plt.ylabel('total')

#plt.savefig('grafico_md_prod_peca_minuto')

plt.show()

```

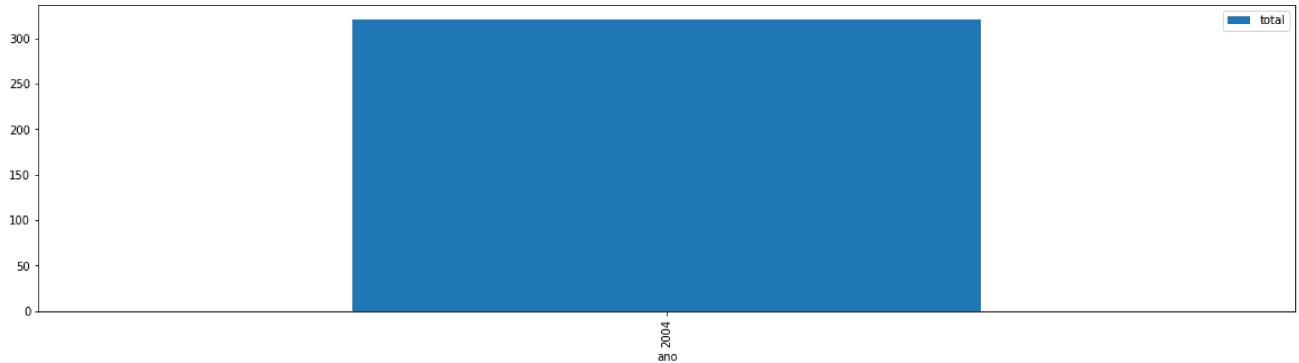


```

dados_empresaStageTotalPorCondominioAnoV2 = dados_empresaStageTotalPorCondominioAno[ [ 'ano' ]
dados_empresaStageTotalPorCondominioAnoV2.plot(figsize=(20,5), kind="bar")

```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa751b6ae90>
```

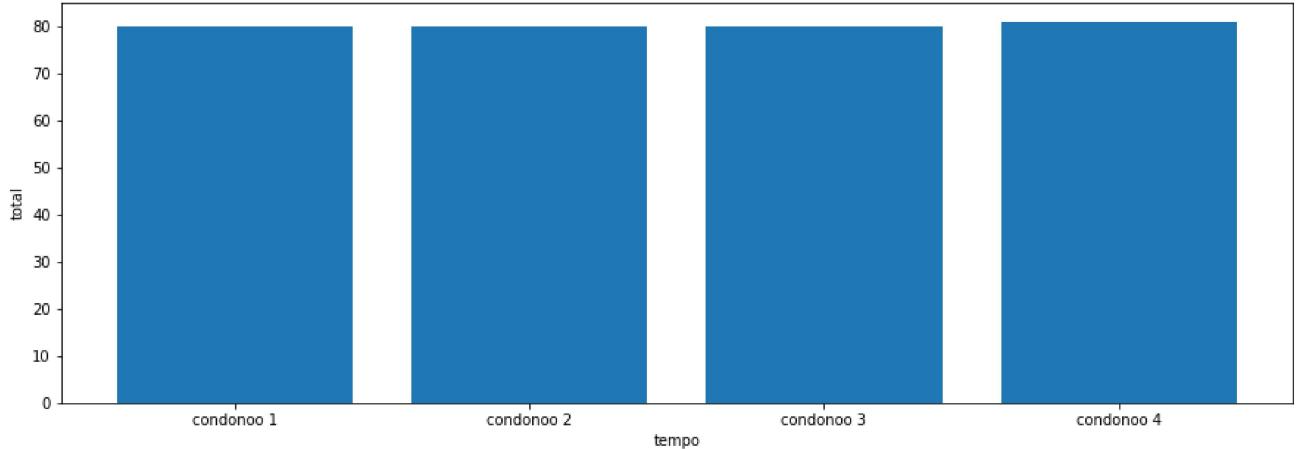


```
#plt.rcParams["figure.figsize"] = (15,5)
plt.bar(dados_empresaStageTotalPorCondominioTotal['NM_FANTASIA'],dados_empresaStageTotalPo
plt.title('Grafico - Total entrada e saida por condominino')
plt.xlabel('tempo')
plt.ylabel('total')

#plt.savefig('grafico_md_prod_peca_minuto')

plt.show()
```

Grafico - Total entrada e saida por condominino

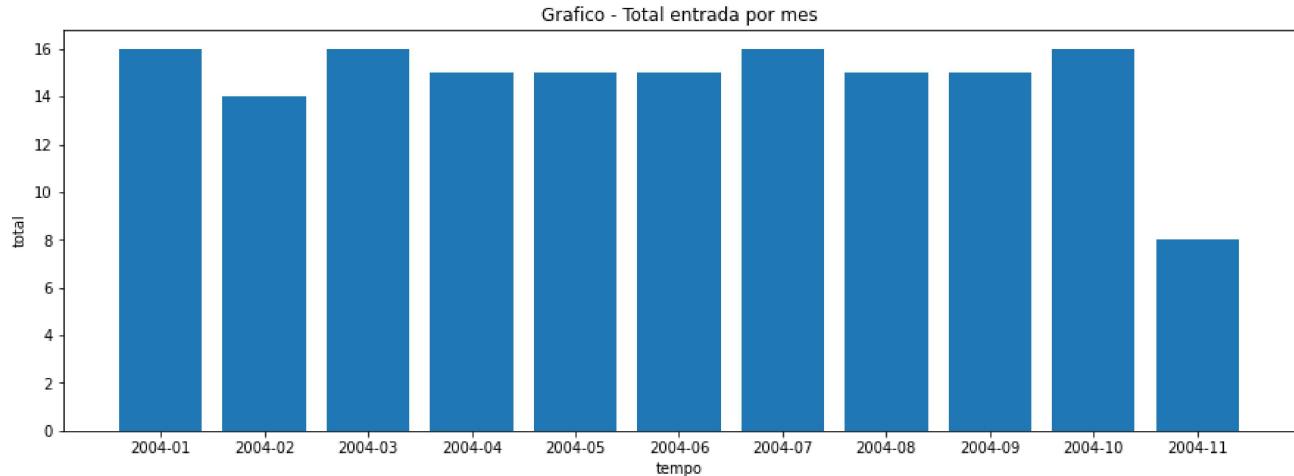


▼ Grafico por entrada

```
#plt.rcParams["figure.figsize"] = (15,5)
plt.bar(dados_empresaStageTotalPorEntradaCondominioPorMes[ 'ano_mes'],dados_empresaStageTot
plt.title('Grafico - Total entrada por mes')
plt.xlabel('tempo')
plt.ylabel('total')
```

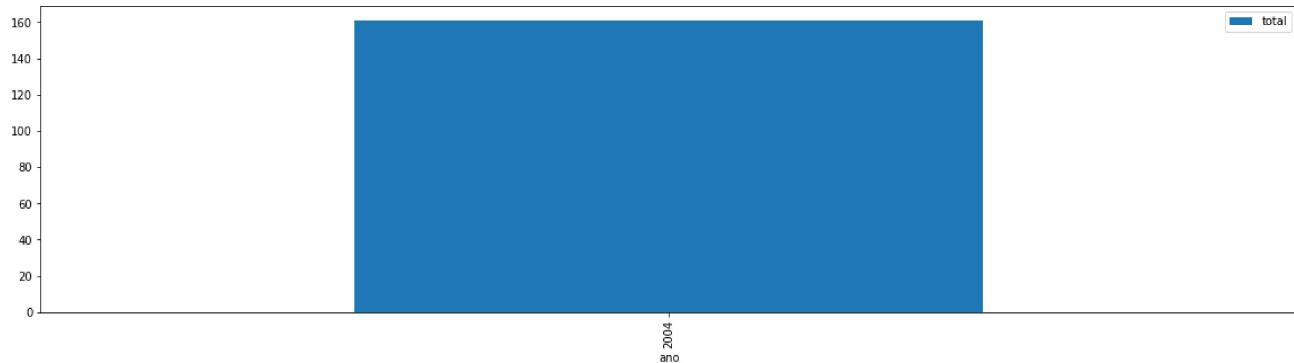
```
#plt.savefig('grafico_md_prod_peca_minuto')
```

```
plt.show()
```



```
dados_empresaStageTotalPorEntradaCondominioAnoV2 = dados_empresaStageTotalPorEntradaCondominioAnoV2.plot(figsize=(20,5), kind="bar")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa7519d0610>
```



```
#plt.rcParams["figure.figsize"] = (15,5)
```

```
plt.bar(dados_empresaStageTotalPorEntradaCondominioTotal['NM_FANTASIA'],dados_empresaStageTotalPorEntradaCondominioTotal['NM_FANTASIA'])
```

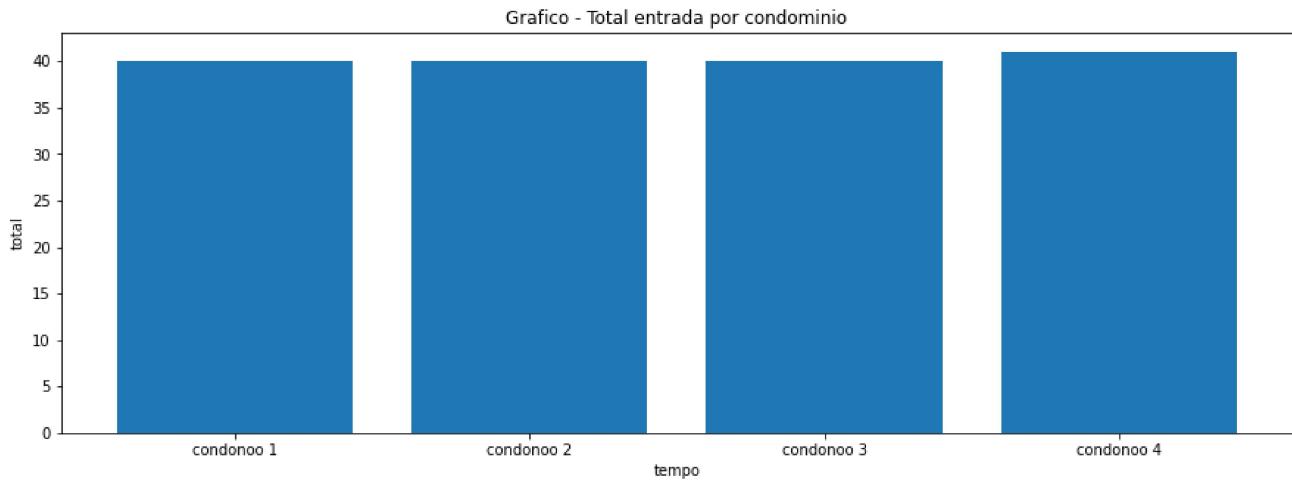
```
plt.title('Grafico - Total entrada por condominio')
```

```
plt.xlabel('tempo')
```

```
plt.ylabel('total')
```

```
#plt.savefig('grafico_md_prod_peca_minuto')
```

```
plt.show()
```

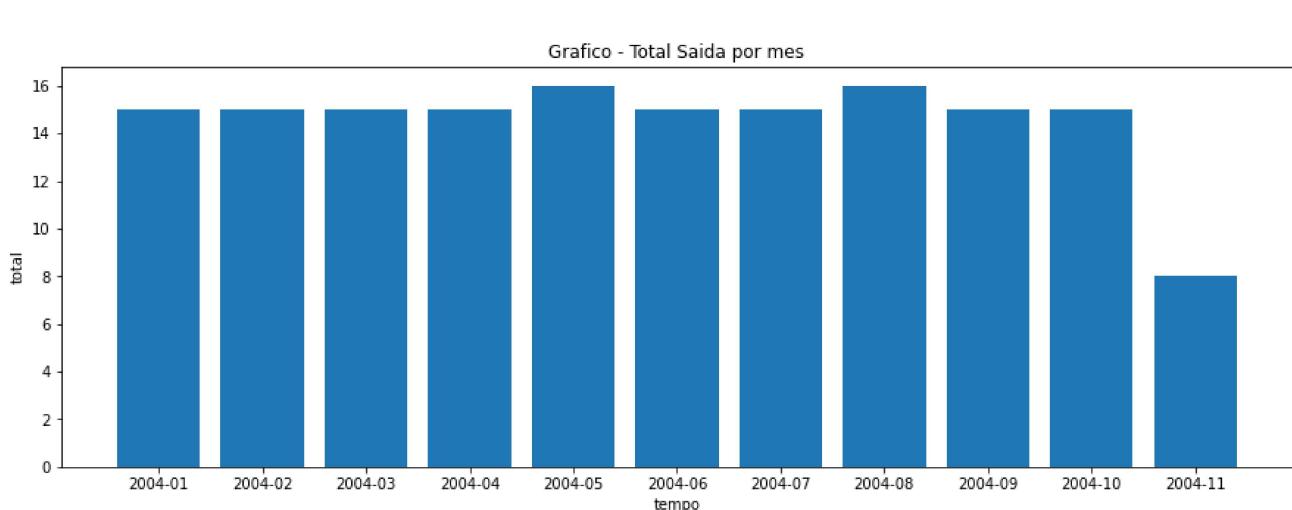


▼ Grafico por saida

```
#plt.rcParams["figure.figsize"] = (15,5)
plt.bar(dados_empresaStageTotalPorSaidaCondominioPorMes[ 'ano_mes' ],dados_empresaStageTotal
plt.title('Grafico - Total Saida por mes')
plt.xlabel('tempo')
plt.ylabel('total')

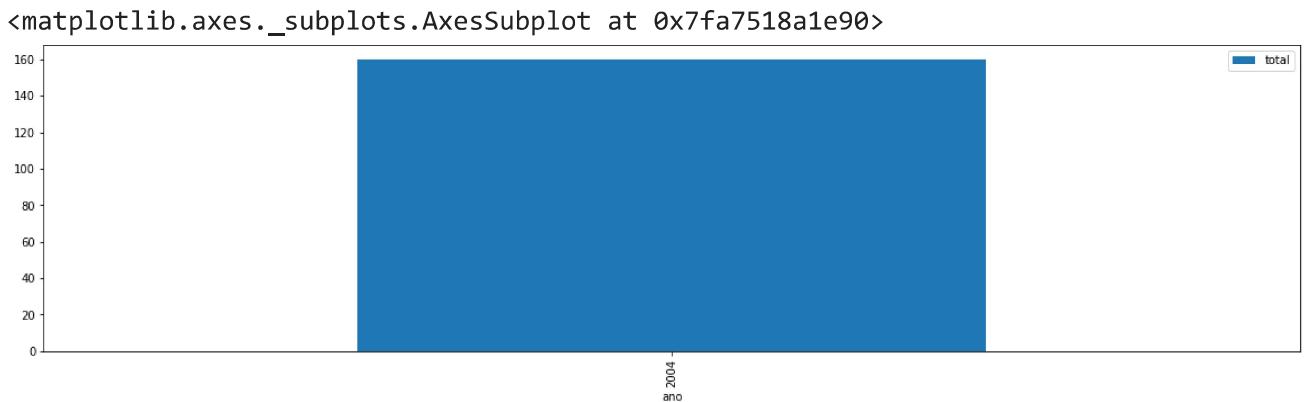
# plt.savefig('grafico_md_prod_peca_minuto')

plt.show()
```



`dados_empresaStageTotalPorSaidaCondominioAnoV2 = dados_empresaStageTotalPorSaidaCondominio.`

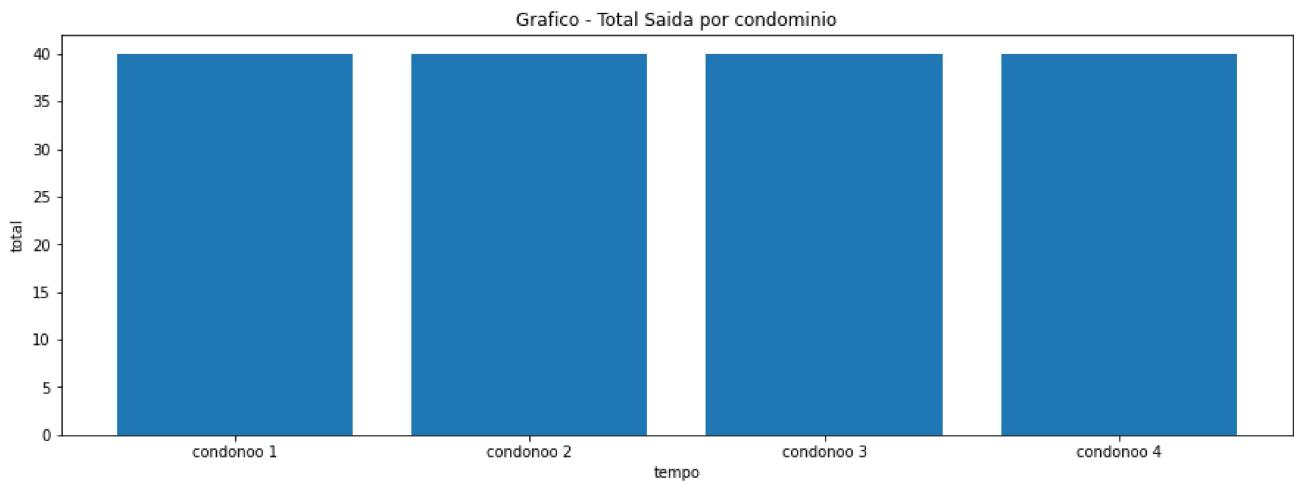
```
dados_empresaStageTotalPorSaidaCondominioAnoV2.plot(figsize=(20,5), kind="bar")
```



```
#plt.rcParams["figure.figsize"] = (15,5)
plt.bar(dados_empresaStageTotalPorSaidaCondominioTotal['NM_FANTASIA'],dados_empresaStageTo
plt.title('Grafico - Total Saida por condominio')
plt.xlabel('tempo')
plt.ylabel('total')

#plt.savefig('grafico_md_prod_peca_minuto')

plt.show()
```

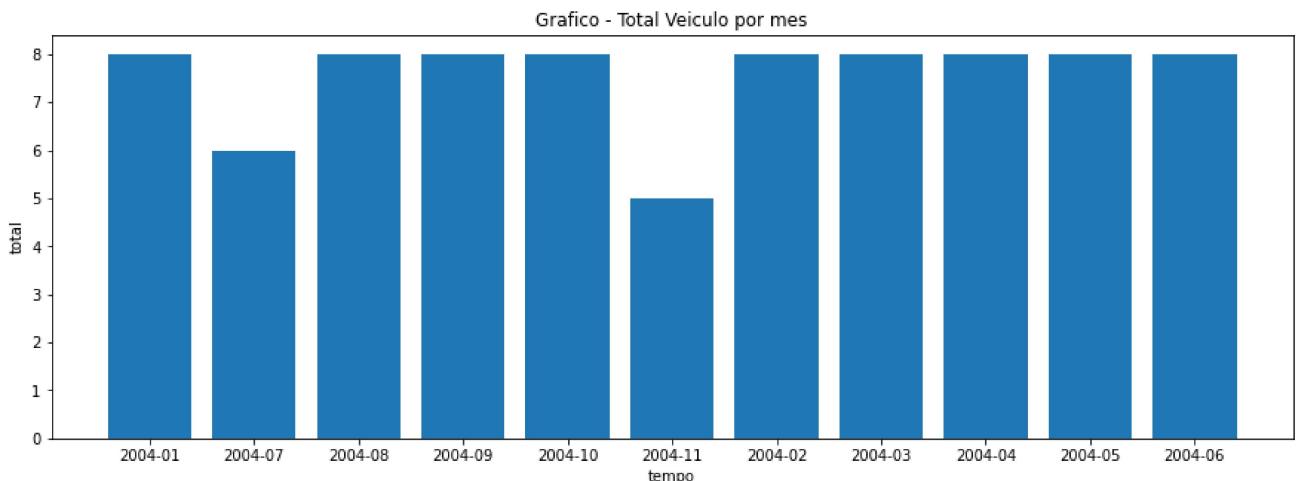


▼ Grafico por veiculo

```
#plt.rcParams["figure.figsize"] = (15,5)
plt.bar(dados_empresaStageTotalPorVeiculoCondominioAnoMes[ 'ano_mes' ],dados_empresaStageTot
plt.title('Grafico - Total Veiculo por mes')
plt.xlabel('tempo')
plt.ylabel('total')

#plt.savefig('grafico_md_prod_peca_minuto')

plt.show()
```



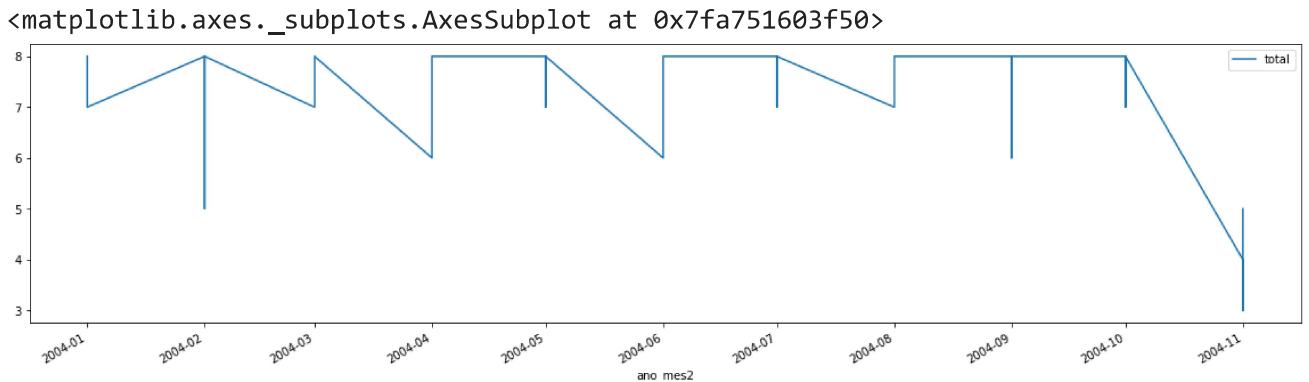
▼ Graficos Serie Temporais

```
# series temporais
from statsmodels.tsa.stattools import adfuller
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
# from statsmodels.tsa.arima_model import ARIMA
from statsmodels.tsa.arima.model import ARIMA
from sklearn.metrics import mean_squared_error, mean_absolute_percentage_error, mean_absol

dados_empresaStageTotalPorCondominioAnoMesV2 = dados_empresaStageTotalPorCondominio[['ano_'

dados_empresaStageTotalPorCondominioAnoMesV2.set_index('ano_mes2', inplace=True)

dados_empresaStageTotalPorCondominioAnoMesV2.plot(figsize=(20,5))
```



▼ Verificar estacionariedade

```
teste_adf = adfuller(dados_empresaStageTotalPorCondominioAnoMesV2['total'])

print('Estatística ADF: {:.4}'.format(teste_adf[0]))

Estatística ADF: -1.72

print('p-value: {:.4}'.format(teste_adf[1]))

p-value: 0.421
```

▼ Modelo ARIMA

▼ Integração I(d)

```
teste_adf = adfuller(dados_empresaStageTotalPorCondominioAnoMesV2['total'].diff().dropna())

print('Estatística ADF: {:.4}'.format(teste_adf[0]))

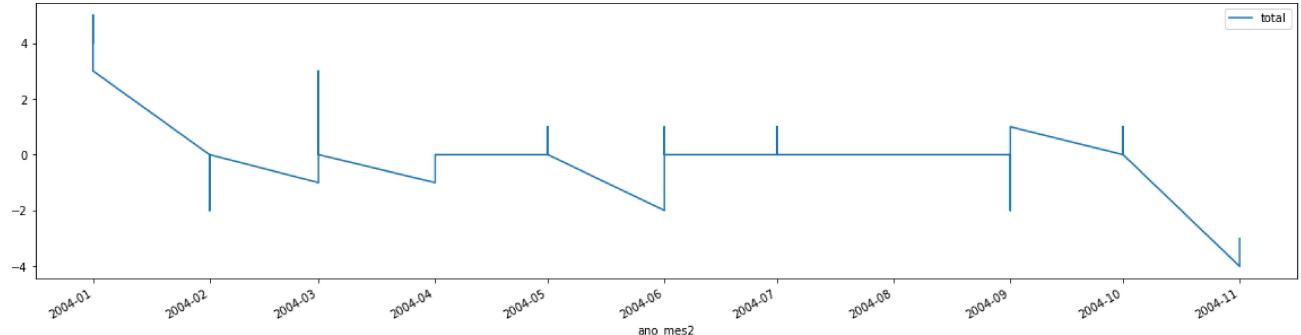
Estatística ADF: -7.004

print('p-value: {:.4}'.format(teste_adf[1]))
# abaixo é gerado p-value maior que 5.0, monstrando que o modelo se tornou estacionario

p-value: 7.197e-10

dados_empresaStageTotalPorCondominioAnoMesV2.diff().plot(figsize=(20,5))
```

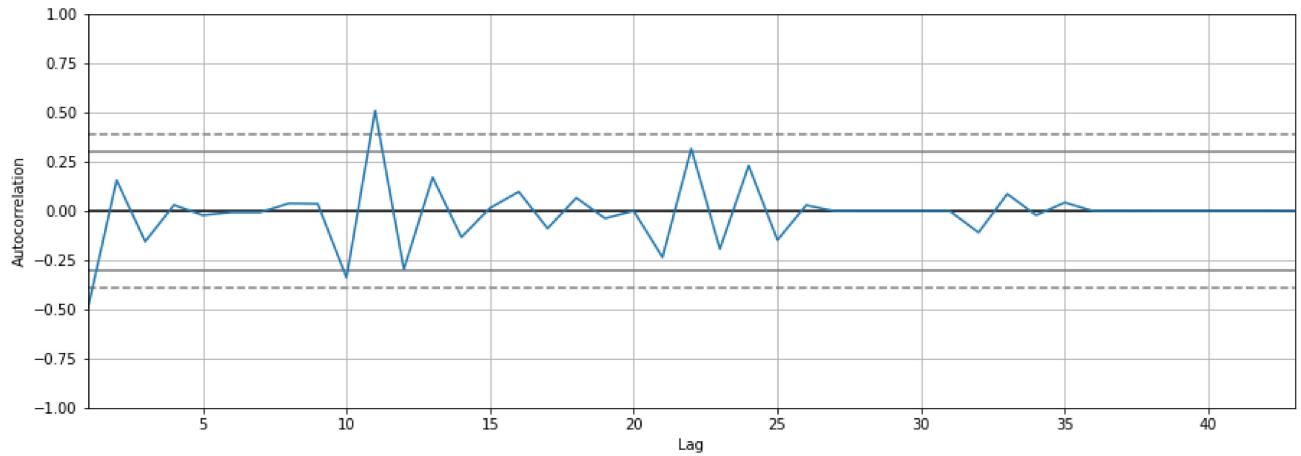
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fa751446e90>
```



▼ Parâmetros AR(p) - analisando a autocorrelação parcial

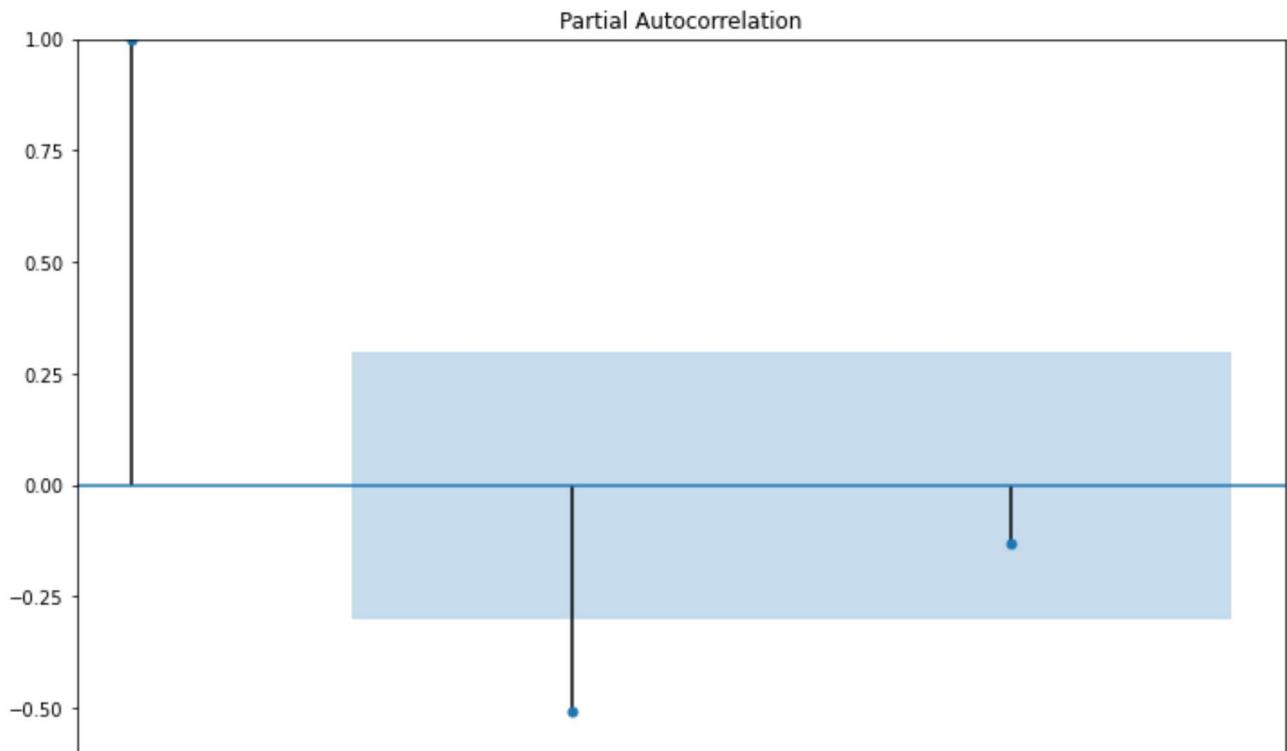
```
from pandas.plotting import autocorrelation_plot
```

```
autocorrelation_plot(dados_empresaStageTotalPorCondominioAnoMesV2['total'].diff().dropna())
plt.show()
```



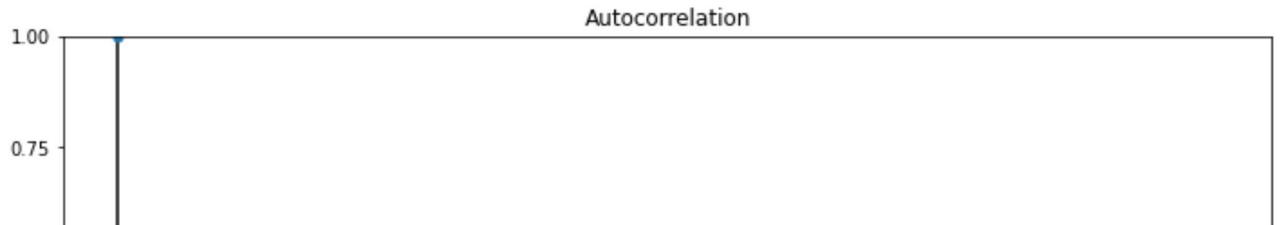
```
fig = plt.figure(figsize=(12,20))
ax1 = fig.add_subplot(211)
fig = plot_pacf(dados_empresaStageTotalPorCondominioAnoMesV2['total'].diff().dropna(), lag
# no modelo abaixo há duas defagem no modelo AR, sendo a maior com valor 10
```

```
/usr/local/lib/python3.7/dist-packages/statsmodels/graphics/tsplots.py:353: FutureWarning,
```



▼ Parâmetros MA(q) analisando a autocorrelação

```
|  
|  
fig = plt.figure(figsize=(12,20))  
ax1 = fig.add_subplot(211)  
fig = plot_acf(dados_empresaStageTotalPorCondominioAnoMesV2['total'].diff().dropna(), lags  
  
# abaixo há dois valores fora da linha tracejada, sendo a maior 11
```



▼ Estimação do modelo ARIMA(p,d,q) - treinando o modelo

considerando 7 dias no futuro

test size = 7

```
dados_treino = dados_empresaStageTotalPorCondominioAnoMesV2[:-test_size]
dados teste = dados_empresaStageTotalPorCondominioAnoMesV2[-test_size:]
```

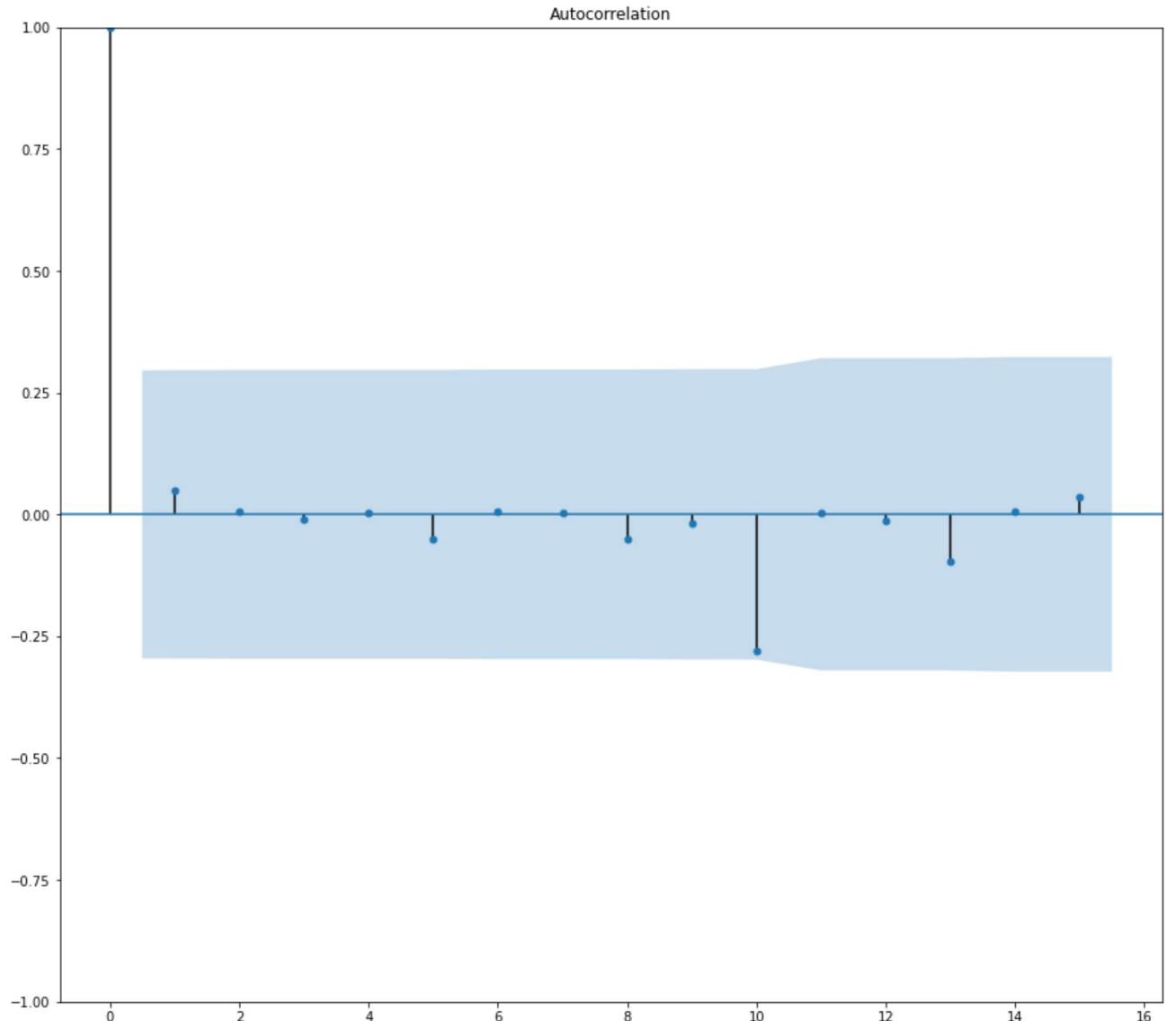
▼ Modelo ARIMA

```
modelo = ARIMA(dados_empresaStageTotalPorCondominioAnoMesV2['total'].values, order=(10, 1, 1))
modelo.summary()
```

SARIMAX Results

```
# acf
fig = plt.figure(figsize=(15,30))
ax1 = fig.add_subplot(211)
fig = plot_acf(modelo.resid, lags=15, ax=ax1)

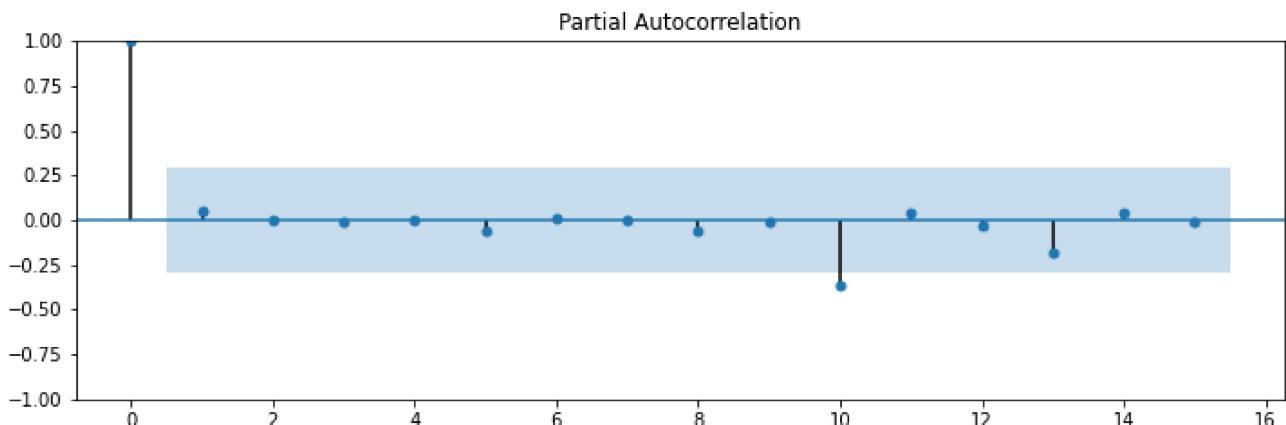
# o grafico abaixo mostra que nenhum ponto está fora da area tracejada, o modelo ficou MA(
```



```
# pacf
fig = plt.figure(figsize=(12,8))
```

```
ax1 = fig.add_subplot(211)
fig = plot_pacf(modelo.resid, lags=15, ax=ax1)
```

```
/usr/local/lib/python3.7/dist-packages/statsmodels/graphics/tsplots.py:353: FutureWarning:
```



Avaliação modelo ARIMA

▼ Dados outliers

```
dados_empresaStageTotalPorCondominioV3 = dados_empresaStageTotalPorCondominio.copy()
dados_empresaStageTotalPorCondominioV3.head()
```

	NM_FANTASIA	ano_mes2	total
0	condonoo 1	2004-01-01	8
1	condonoo 1	2004-02-01	8
2	condonoo 1	2004-03-01	8
3	condonoo 1	2004-04-01	8
4	condonoo 1	2004-05-01	8

```
from sklearn.preprocessing import scale
```

```
dados_empresaStageTotalPorCondominioV3['total_outliers'] = scale(dados_empresaStageTotalPo
```

```
dados_empresaStageTotalPorCondominioV3[['total_outliers', 'total', 'NM_FANTASIA']].sort_va
```

```
dados_empresaStageTotalPorCondominioV3[['total', 'total_outliers']].boxplot()
```



```
<matplotlib.axes._subplots.AxesSubplot at 0x7f3aa9220f50>
```



```
dados_empresaStageTotalPorCondominioV3.plot.scatter(x = 'total_outliers', y='NM_FANTASIA',
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f3aa93151d0>
```

