

— _v.2

April 30, 2023

1

```
[41]: import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
import warnings
from statsmodels.tsa.statespace.sarimax import SARIMAX
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.stattools import adfuller
from statsmodels.tsa.seasonal import STL
import statsmodels.api as sm
from IPython.display import Markdown
import time
warnings.filterwarnings('ignore')
```

2

csv.

DataFrame.

```
[7]: df = pd.read_csv('StockX-Data-Contest-2019-3.csv')
```

```
[8]: df.head()
```

```
[8]:  Order Date  Brand  Sneaker Name \
0    9/1/17  Yeezy  Adidas-Yeezy-Boost-350-Low-V2-Beluga
1    9/1/17  Yeezy  Adidas-Yeezy-Boost-350-V2-Core-Black-Copper
2    9/1/17  Yeezy  Adidas-Yeezy-Boost-350-V2-Core-Black-Green
3    9/1/17  Yeezy  Adidas-Yeezy-Boost-350-V2-Core-Black-Red
4    9/1/17  Yeezy  Adidas-Yeezy-Boost-350-V2-Core-Black-Red-2017

Sale Price Retail Price Release Date  Shoe Size  Buyer Region
0    $1,097      $220      9/24/16      11.0    California
1     $685      $220     11/23/16      11.0    California
```

2	\$690	\$220	11/23/16	11.0	California
3	\$1,075	\$220	11/23/16	11.5	Kentucky
4	\$828	\$220	2/11/17	11.0	Rhode Island

```
[9]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99956 entries, 0 to 99955
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Order Date      99956 non-null  object
1   Brand           99956 non-null  object
2   Sneaker Name    99956 non-null  object
3   Sale Price      99956 non-null  object
4   Retail Price    99956 non-null  object
5   Release Date    99956 non-null  object
6   Shoe Size       99956 non-null  float64
7   Buyer Region    99956 non-null  object
dtypes: float64(1), object(7)
memory usage: 6.1+ MB
```

```
[10]: df['Brand'] = df['Brand'].str.strip()
```

```
info(
    ,
    (
    datetime)
```

```
[11]: df['Order Date'] = pd.to_datetime(df['Order Date'])
df['Release Date'] = pd.to_datetime(df['Release Date'])
```

```
head(
    )
(object)
,
(int)
```

```
[12]: df['Sale Price'] = df['Sale Price'].str.replace('$', '', regex=True)
df['Sale Price'] = df['Sale Price'].str.replace(',', '', regex=True)
df['Sale Price'] = df['Sale Price'].astype('int')
```

```
[13]: df['Retail Price'] = df['Retail Price'].str.replace('$', '', regex=True)
df['Retail Price'] = df['Retail Price'].str.replace(',', '', regex=True)
df['Retail Price'] = df['Retail Price'].astype('int')
```

```
[14]: df[''] = df['Sale Price'] - df['Retail Price']
```

```
,
data frame,
```

```
[15]: df[df['Buyer Region'].isin(['Wyoming', 'Montana'])]
```

```
[15]:
```

	Order Date	Brand	Sneaker Name	\
597	2017-09-26	Yeezy	Adidas-Yeezy-Boost-350-V2-Cream-White	
3863	2017-11-25	Yeezy	Adidas-Yeezy-Boost-350-V2-Beluga-2pt0	
5854	2017-11-30	Yeezy	Adidas-Yeezy-Boost-350-V2-Cream-White	
8059	2017-12-09	Yeezy	Adidas-Yeezy-Boost-350-V2-Semi-Frozen-Yellow	
8671	2017-12-13	Yeezy	Adidas-Yeezy-Boost-350-V2-Beluga-2pt0	
...	
94443	2019-01-31	Yeezy	Adidas-Yeezy-Boost-350-V2-Zebra	
94827	2019-02-01	Yeezy	adidas-Yeezy-Boost-350-V2-Static	
95660	2019-02-05	Yeezy	Adidas-Yeezy-Boost-350-V2-Semi-Frozen-Yellow	
96836	2019-02-07	Yeezy	Adidas-Yeezy-Boost-350-V2-Blue-Tint	
98670	2019-02-10	Yeezy	adidas-Yeezy-Boost-350-V2-Butter	

	Sale Price	Retail Price	Release Date	Shoe Size	Buyer Region	
597	478	220	2017-04-29	10.5	Wyoming	258
3863	455	220	2017-11-25	11.0	Wyoming	235
5854	450	220	2017-04-29	12.0	Montana	230
8059	670	220	2017-11-18	6.5	Montana	450
8671	352	220	2017-11-25	14.0	Montana	132
...	
94443	337	220	2017-02-25	12.0	Montana	117
94827	272	220	2018-12-27	12.0	Wyoming	52
95660	303	220	2017-11-18	5.5	Wyoming	83
96836	350	220	2017-12-16	13.0	Wyoming	130
98670	230	220	2018-06-30	9.0	Wyoming	10

[89 rows x 9 columns]

2 " , " "

```
[16]: df.groupby(['Order Date', 'Buyer Region'], as_index=False).count()
```

```
[16]:
```

	Order Date	Buyer Region	Brand	Sneaker Name	Sale Price	Retail Price	\
0	2017-09-01	California	6	6	6	6	
1	2017-09-01	Florida	3	3	3	3	
2	2017-09-01	Kansas	1	1	1	1	
3	2017-09-01	Kentucky	1	1	1	1	
4	2017-09-01	Michigan	3	3	3	3	
...	
15066	2019-02-13	Tennessee	3	3	3	3	
15067	2019-02-13	Texas	17	17	17	17	
15068	2019-02-13	Virginia	4	4	4	4	
15069	2019-02-13	Washington	3	3	3	3	
15070	2019-02-13	Wisconsin	2	2	2	2	

	Release Date	Shoe Size	
0	6	6	6
1	3	3	3

2	1	1	1
3	1	1	1
4	3	3	3
...
15066	3	3	3
15067	17	17	17
15068	4	4	4
15069	3	3	3
15070	2	2	2

[15071 rows x 9 columns]

```
[17]: df.groupby(['Order Date', 'Buyer Region']).count()
```

```
[17]:
```

		Brand	Sneaker Name	Sale Price	Retail Price	\
Order Date	Buyer Region					
2017-09-01	California	6	6	6	6	
	Florida	3	3	3	3	
	Kansas	1	1	1	1	
	Kentucky	1	1	1	1	
	Michigan	3	3	3	3	
...		
2019-02-13	Tennessee	3	3	3	3	
	Texas	17	17	17	17	
	Virginia	4	4	4	4	
	Washington	3	3	3	3	
	Wisconsin	2	2	2	2	

		Release Date	Shoe Size	
Order Date	Buyer Region			
2017-09-01	California	6	6	6
	Florida	3	3	3
	Kansas	1	1	1
	Kentucky	1	1	1
	Michigan	3	3	3
...	
2019-02-13	Tennessee	3	3	3
	Texas	17	17	17
	Virginia	4	4	4
	Washington	3	3	3
	Wisconsin	2	2	2

[15071 rows x 7 columns]

```
[91]: def brand_sezon(model_data, model):
        modell_data = model_data.groupby("Sale Price", as_index = False).count()
        x_data = modell_data['Sale Price']
```

```

y_data = modell_data['Order Date']
fig = plt.figure(figsize = (20,5))
plt.title(f"                                {model}")
plt.xticks(rotation=90)
plt.scatter(x_data, y_data)

# modell_data = model_data.groupby("Order Date", as_index = False).count()
x_data = model_data['Order Date']
y_data = model_data['Sale Price']
fig = plt.figure(figsize = (20,5))
plt.title(f"                                {model}")
plt.xticks(rotation=90)
plt.scatter(x_data, y_data, color='g')

def print_sale(brand_data, title):
    '''                                '''
    month_sale = brand_data.groupby('Order Month', as_index = False).count()
    x_data = month_sale['Order Month']
    y_data = month_sale['Order Date']
    fig = plt.figure(figsize = (20,5))
    plt.title(title)
    plt.xticks(rotation=90)
    plt.bar(x_data, y_data)

def print_all_sale(data, title):
    brand_count = data.groupby("Brand", as_index = False).count()
    x_data = brand_count['Brand']
    y_data = brand_count['Order Date']
    fig = plt.figure(figsize = (20,5))
    plt.title(title)
    plt.xticks(rotation=90)
    plt.bar(x_data, y_data)

def print_heatmap(data, x, y, z, annotation = True):
    sns.set()
    fig, ax = plt.subplots(figsize = (20, 20))
    heatmap = sns.heatmap(data = data.pivot(x, y, z), ax = ax, cmap = 'plasma',
↪annot = annotation)
    plt.show()

def sarima(brand_data):
    '''                                '''

```

```

sales_train_row = brand_data.groupby('Order Date', as_index=False).
↳agg({'Sneaker Name': 'count'})
sales_train_row.rename(columns = {'Sneaker Name' : 'Retail Count'}, inplace_
↳= True)
sales_train_row.set_index(sales_train_row['Order Date'], inplace=True)
sales_index = pd.date_range(start=sales_train_row['Order Date'].index.
↳min(), end=sales_train_row['Order Date'].index.max(), freq='D')
sales_ts = pd.Series(index=sales_index)
sales_ts = sales_ts.combine_first(sales_train_row['Retail Count'])
sales_ts = sales_ts.resample('W').sum()
#
fig = plt.figure(figsize = (20,5)) #
plt.xticks(rotation=90) # x
plt.title(' ')
plt.plot(sales_ts)
#
stl = STL(sales_ts, seasonal=13)
res = stl.fit()
fig = plt.figure(figsize = (20,5)) #
plt.xticks(rotation=90) # x
plt.title(' ')
plt.plot(res.trend)
fig = plt.figure(figsize = (20,5)) #
plt.xticks(rotation=90) # x
plt.title(' ')
plt.plot(res.seasonal)
# ACF PACF
fig, axes = plt.subplots(2, 1, figsize=(10, 8))
plt.title(' ACF PACF')
plot_acf(sales_ts, ax=axes[0])
plot_pacf(sales_ts, ax=axes[1])
plt.show()
#
display(Markdown('### '))
ADF_result = adfuller(sales_ts)
print(f'ADF :{ADF_result[0]}')
print(f'p-value: {ADF_result[1]}')
print(' :')
for key, value in ADF_result[4].items():
    print(key, value)
# SARIMA
p = 1
d = 1
q = 1
P = 1
D = 1
Q = 1

```

```

s = 27
model = SARIMAX(sales_ts, order=(p, d, q), seasonal_order=(P, D, Q, s))
results = model.fit()
forecast = results.predict(start=len(sales_ts), end=len(sales_ts)+54,
dynamic=True)
plt.figure(figsize=(10, 6))
plt.plot(sales_ts.index, sales_ts.values, label='Sales')
plt.plot(sales_ts.index, sales_ts.values, label='Sales')
plt.plot(forecast.index, forecast.values, label='Forecast')
plt.fill_between(forecast.index, forecast.quantile(0), forecast.
quantile(1), alpha=0.2, label='Forecast')
plt.title('Sales and Forecast')
plt.xlabel('Date')
plt.ylabel('Sales')
plt.legend()
plt.show()

def analitica_region(region, data):
display(Markdown(f'# Sales by Region: {region}'))
#
data = data[data['Buyer Region'].isin([region])]
#
brands = data['Brand'].unique()
#
data['Order Month'] = data['Order Date'].dt.month

#
for brand in brands:
markdown_text = f'# Sales by Brand: {brand}'
display(Markdown(markdown_text))
#
brand_data = data[data['Brand'].isin([brand])]
#
print_sale(brand_data, f'Sales by Brand: {brand}')
#
models = brand_data['Sneaker Name'].unique()
#
count_sales = len(brand_data)
#
sarima(brand_data)

```

```

count_brand = brand_data.groupby(['Order Month', 'Sneaker Name'],
as_index = False).agg({'Sale Price' : 'count', ' ' : 'sum'})
size_model = brand_data.groupby(['Shoe Size', 'Sneaker Name'], as_index_
⇒ False).agg({'Sale Price' : 'count', ' ' : 'sum'})
# index_max = month_sale['Sale Price'].idxmax()
# display(month_sale.loc[index_max]['Order Month'])
#
display(Markdown('### '))
for model in models:
    #
    model_data = brand_data[brand_data['Sneaker Name'].isin([model])]
    if len(model_data['Order Month'].unique()) == 12:
        print(f' {brand} - {model}')
        brand_sezon(model_data, model)
    else:
        print(f' {brand} - {model}')

print(f' {brand} ')
print_heatmap(count_brand, 'Sneaker Name', 'Order Month', "Sale Price")
print(f' {brand} ')
print_heatmap(count_brand, 'Sneaker Name', 'Order Month', " ")
print(f' {brand} ')
print_heatmap(size_model, 'Sneaker Name', 'Shoe Size', 'Sale Price')
print(f' {brand} ')
print_heatmap(size_model, 'Sneaker Name', 'Shoe Size', ' ', annotation_
⇒ False)

#
brand_str = ", ".join(brands)
print_all_sale(data, f" {brand_str}")

```

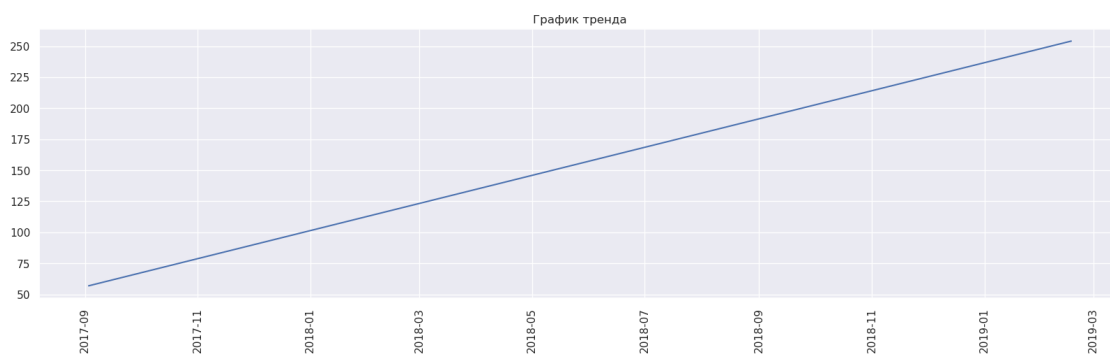
```
[90]: analitica_region("California", df)
```

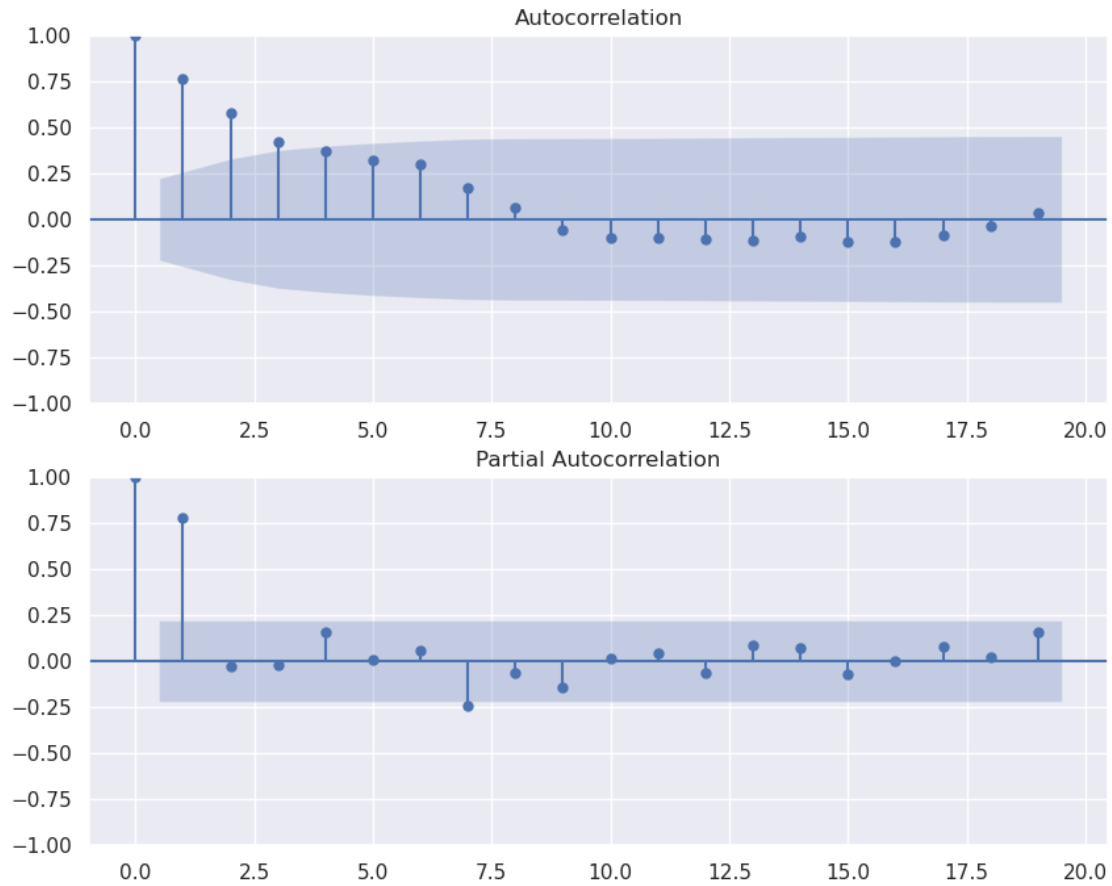
3

4

Yeezy







4.0.1

ADF : -3.1761944774875244

p-value: 0.021404653667943998

:

1% -3.5194805351545413

5% -2.9003945086747343

10% -2.5874984279778395

RUNNING THE L-BFGS-B CODE

* * *

Machine precision = 2.220D-16

N = 5 M = 10

At X0 0 variables are exactly at the bounds

At iterate 0 f= 4.00198D+00 |proj g|= 1.03075D-01

This problem is unconstrained.

At iterate	5	f=	3.98161D+00	proj g =	5.21194D-03
At iterate	10	f=	3.97982D+00	proj g =	1.74104D-03
At iterate	15	f=	3.97706D+00	proj g =	1.54426D-02
At iterate	20	f=	3.96672D+00	proj g =	3.64833D-04
At iterate	25	f=	3.96671D+00	proj g =	8.91724D-04
At iterate	30	f=	3.96667D+00	proj g =	1.02909D-03
At iterate	35	f=	3.96662D+00	proj g =	2.48054D-03
At iterate	40	f=	3.96658D+00	proj g =	1.74391D-03
At iterate	45	f=	3.96651D+00	proj g =	9.32063D-04
At iterate	50	f=	3.96649D+00	proj g =	5.03136D-04

* * *

Tit = total number of iterations
 Tnf = total number of function evaluations
 Tnint = total number of segments explored during Cauchy searches
 Skip = number of BFGS updates skipped
 Nact = number of active bounds at final generalized Cauchy point
 Projg = norm of the final projected gradient
 F = final function value

* * *

N	Tit	Tnf	Tnint	Skip	Nact	Projg	F
5	50	62	1	0	0	5.031D-04	3.966D+00
F =	3.9664907594093148						

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT



4.0.2

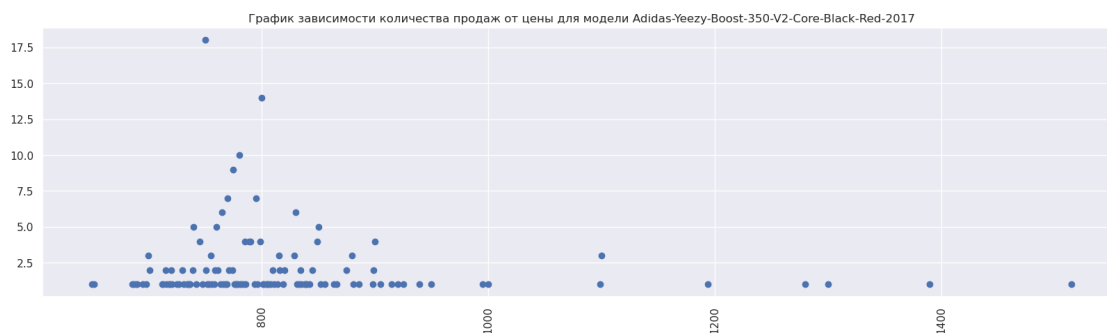
Yeezy	-	Adidas-Yeezy-
Boost-350-Low-V2-Beluga	-	Adidas-Yeezy-
Yeezy	-	Adidas-Yeezy-
Boost-350-V2-Core-Black-Copper	-	Adidas-Yeezy-Boost-350-V2-Core-
Black-Green	-	Adidas-Yeezy-Boost-350-V2-Core-
Yeezy	-	Adidas-Yeezy-Boost-350-V2-Zebra
Black-White	-	Adidas-Yeezy-Boost-350-V2-Cream-
Yeezy	-	Adidas-Yeezy-Boost-350-V2-Core-
Black-Red-2017	-	Adidas-Yeezy-
Yeezy	-	Adidas-Yeezy-Boost-350-Low-
Boost-350-V2-Core-Black-Red	-	Adidas-Yeezy-Boost-350-Low-
Yeezy	-	Adidas-Yeezy-Boost-350-Low-
Turtledove	-	Adidas-Yeezy-Boost-350-Low-
Yeezy	-	Adidas-Yeezy-Boost-350-Low-
Moonrock	-	Adidas-Yeezy-Boost-350-Low-
Yeezy	-	Adidas-Yeezy-Boost-350-Low-
Pirate-Black-2015	-	Adidas-Yeezy-Boost-350-Low-
Yeezy	-	Adidas-Yeezy-Boost-350-Low-
Pirate-Black-2016	-	Adidas-Yeezy-Boost-350-Low-

Yeezy	-	Adidas-Yeezy-Boost-350-V2-Semi-
Frozen-Yellow		
Yeezy	-	Adidas-Yeezy-
Boost-350-V2-Beluga-2pt0		
Yeezy	-	Adidas-Yeezy-Boost-350-Low-
Oxford-Tan		
Yeezy	-	Adidas-Yeezy-Boost-350-V2-Blue-
Tint		
Yeezy	-	adidas-Yeezy-
Boost-350-V2-Butter		
Yeezy	-	Adidas-Yeezy-
Boost-350-V2-Sesame		
Yeezy	-	adidas-Yeezy-
Boost-350-V2-Static		
Yeezy	-	adidas-Yeezy-
Boost-350-V2-Static-Reflective		
Yeezy		

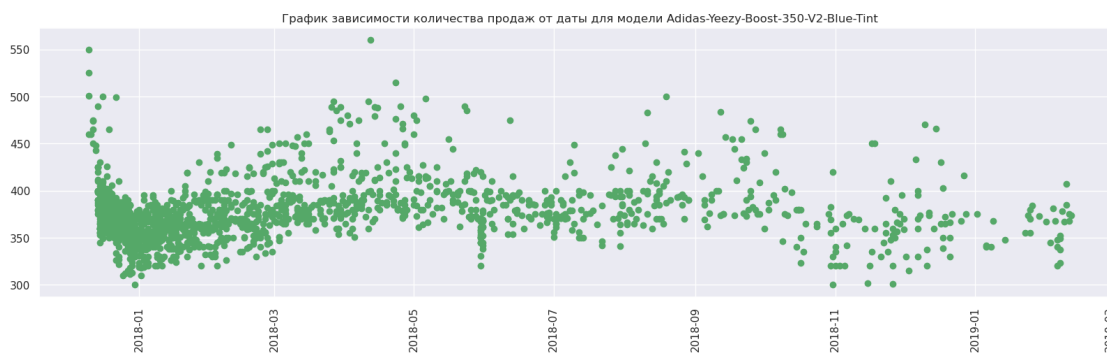


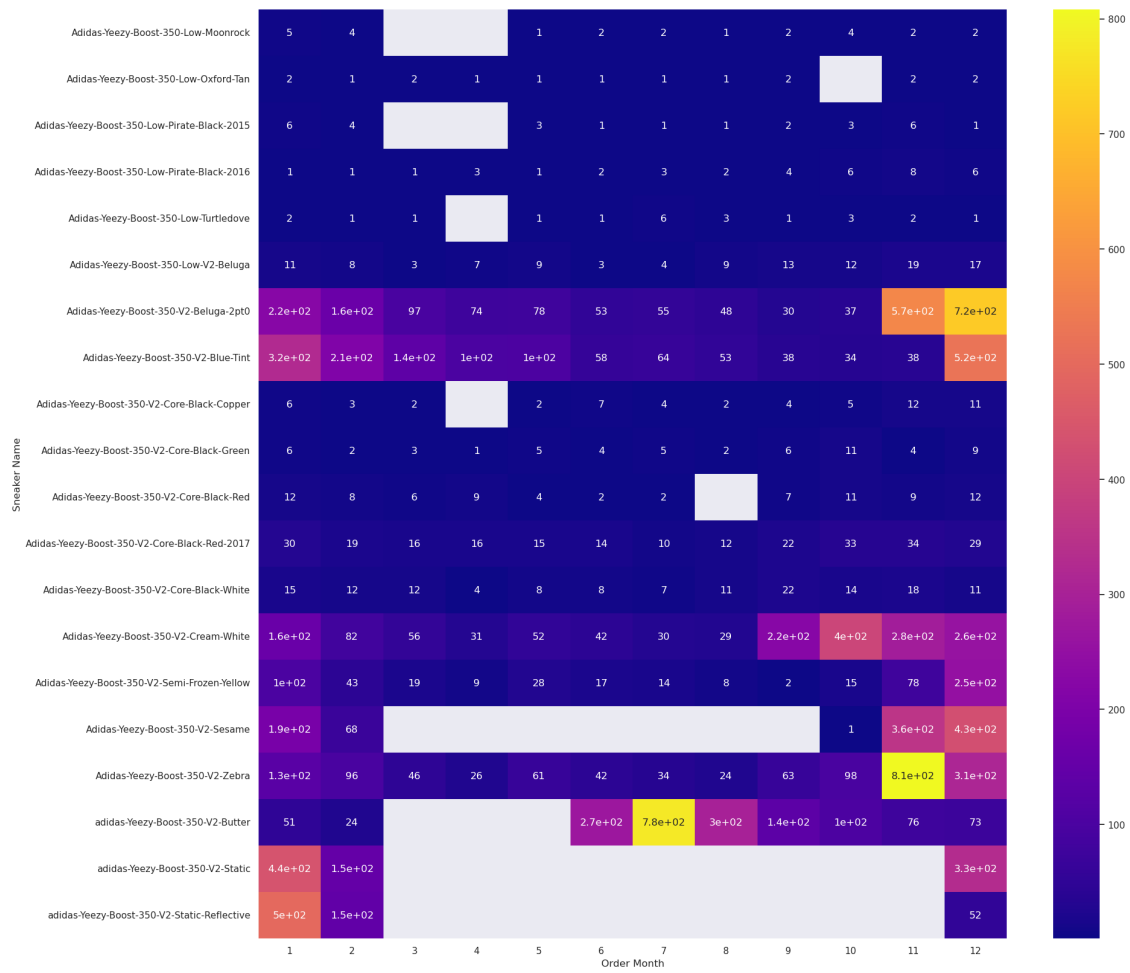








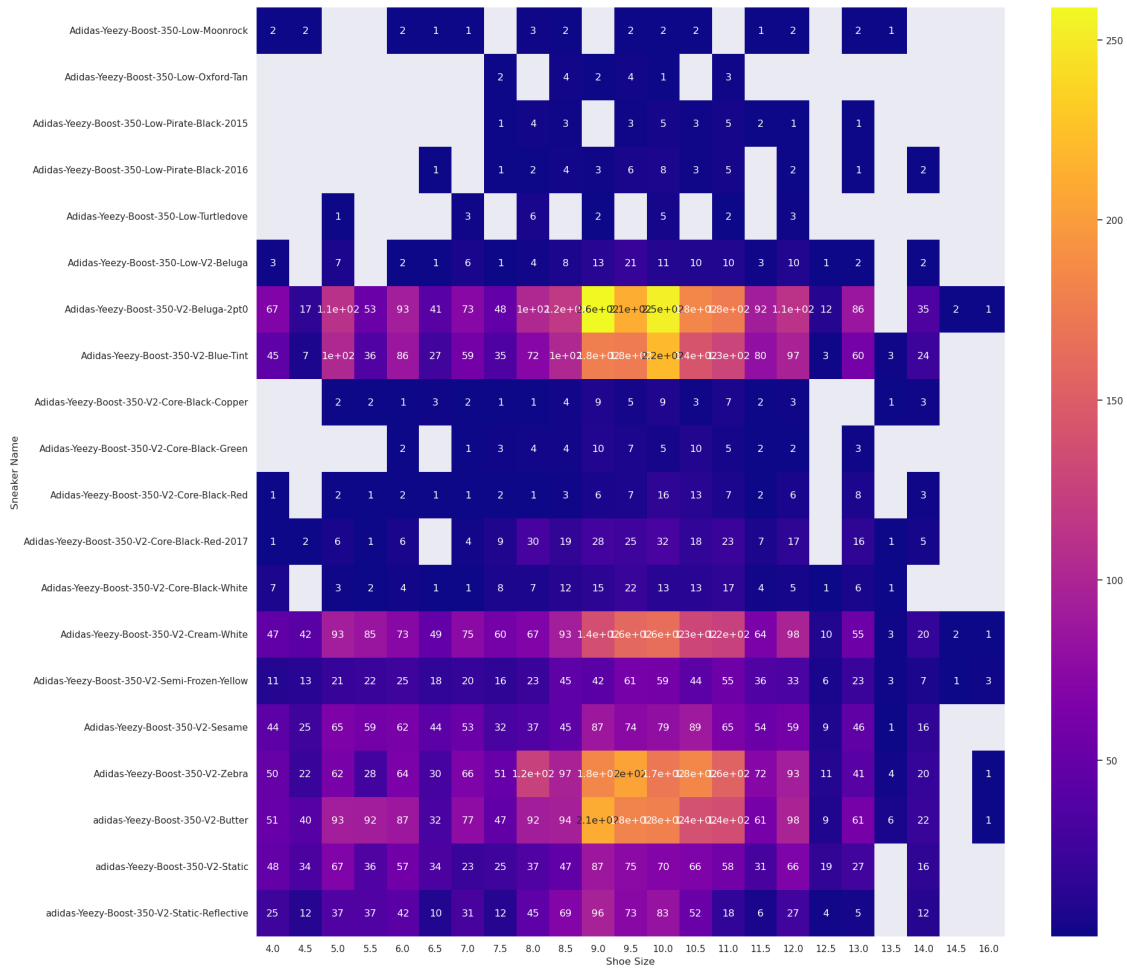




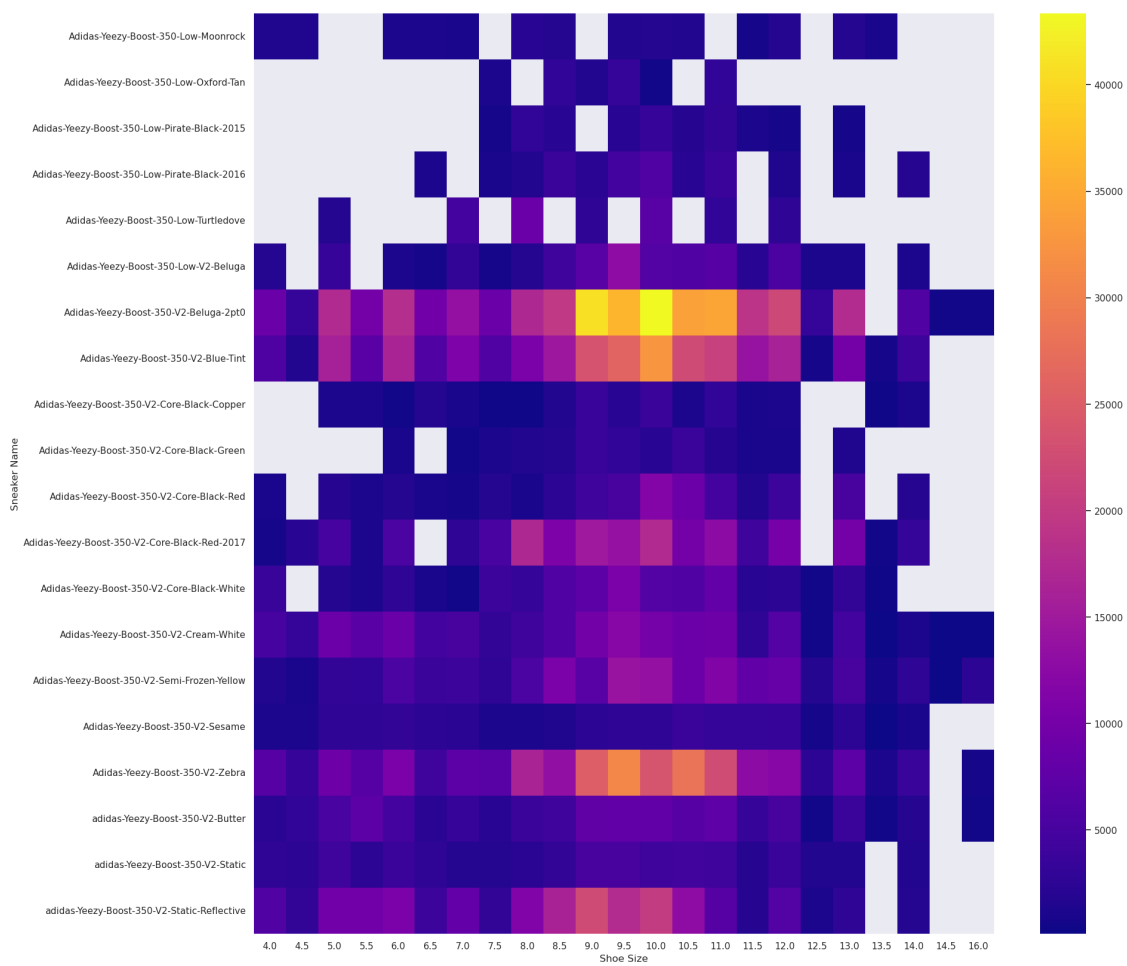
Yeezy



Yeezy

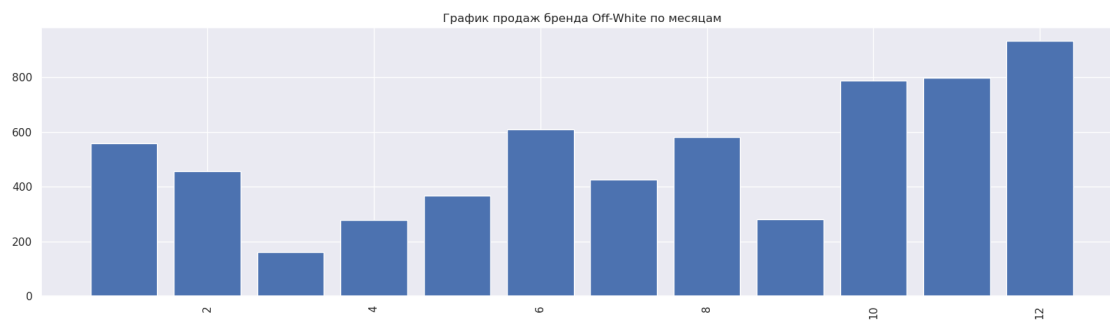


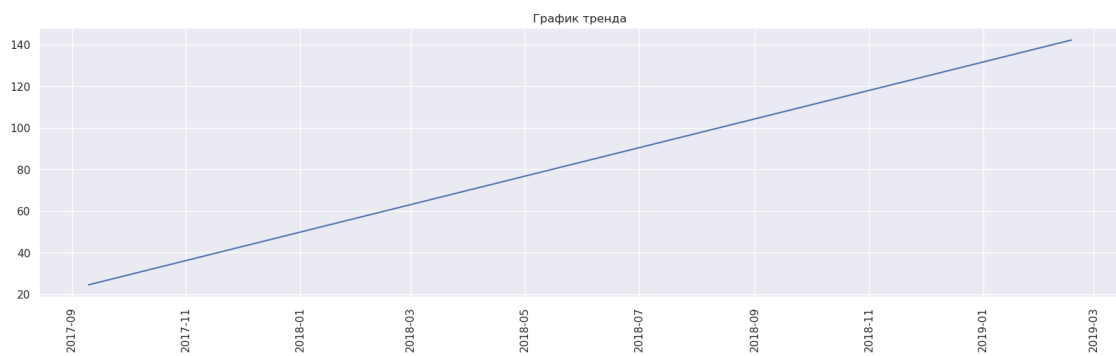
Yeezy

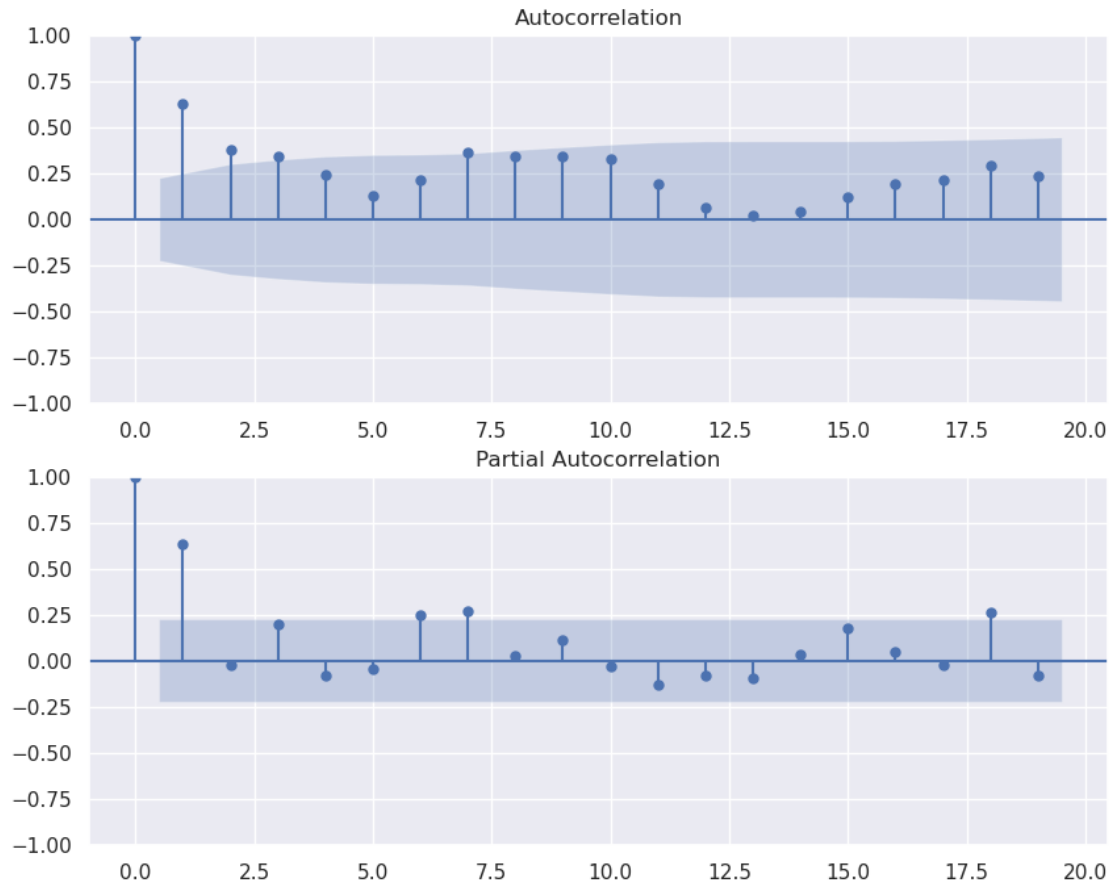


5

Off-White







5.0.1

ADF : -1.8397153931046681

p-value: 0.3609298773712718

:

1% -3.528889992207215

5% -2.9044395987933362

10% -2.589655654274312

RUNNING THE L-BFGS-B CODE

* * *

Machine precision = 2.220D-16

N = 5 M = 10

At X0 0 variables are exactly at the bounds

At iterate 0 f= 3.47638D+00 |proj g|= 9.95138D-02

This problem is unconstrained.

At iterate	5	f=	3.44120D+00	proj g =	6.86528D-03
At iterate	10	f=	3.43457D+00	proj g =	1.50240D-02
At iterate	15	f=	3.41830D+00	proj g =	8.67628D-04
At iterate	20	f=	3.41743D+00	proj g =	1.96293D-02
At iterate	25	f=	3.41491D+00	proj g =	1.51508D-04
At iterate	30	f=	3.41491D+00	proj g =	4.14626D-04
At iterate	35	f=	3.41490D+00	proj g =	4.02367D-04
At iterate	40	f=	3.41490D+00	proj g =	3.83251D-04
At iterate	45	f=	3.41489D+00	proj g =	1.33798D-04
At iterate	50	f=	3.41489D+00	proj g =	3.42082D-04

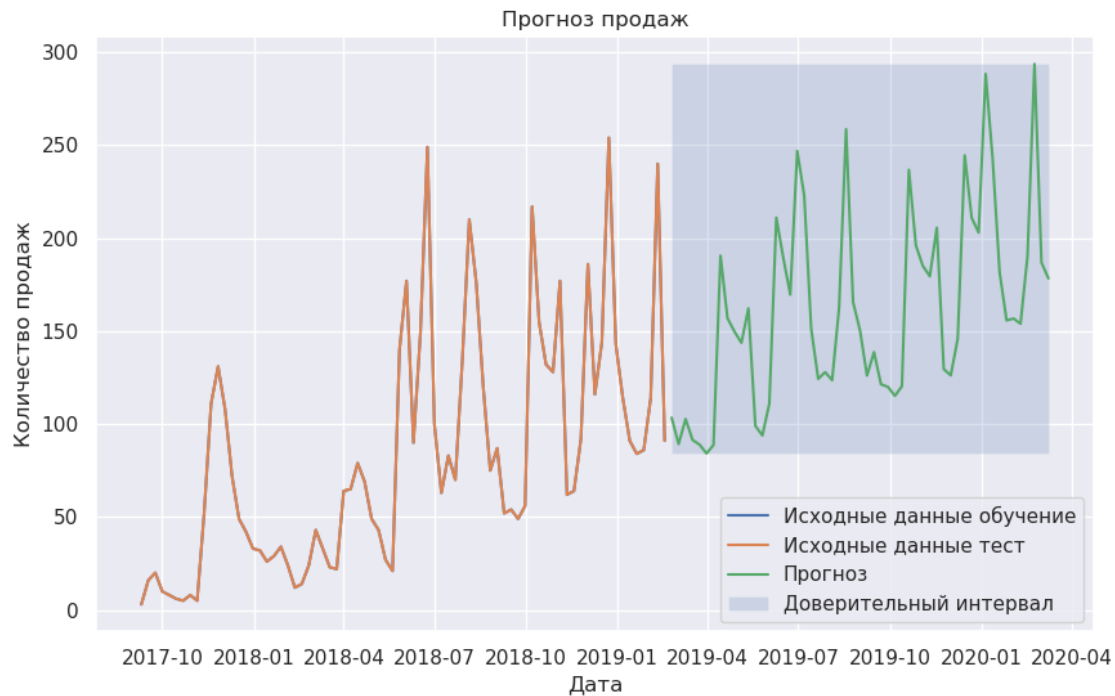
* * *

Tit = total number of iterations
 Tnf = total number of function evaluations
 Tnint = total number of segments explored during Cauchy searches
 Skip = number of BFGS updates skipped
 Nact = number of active bounds at final generalized Cauchy point
 Projg = norm of the final projected gradient
 F = final function value

* * *

N	Tit	Tnf	Tnint	Skip	Nact	Projg	F
5	50	66	1	0	0	3.421D-04	3.415D+00
F =	3.4148854465621774						

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT



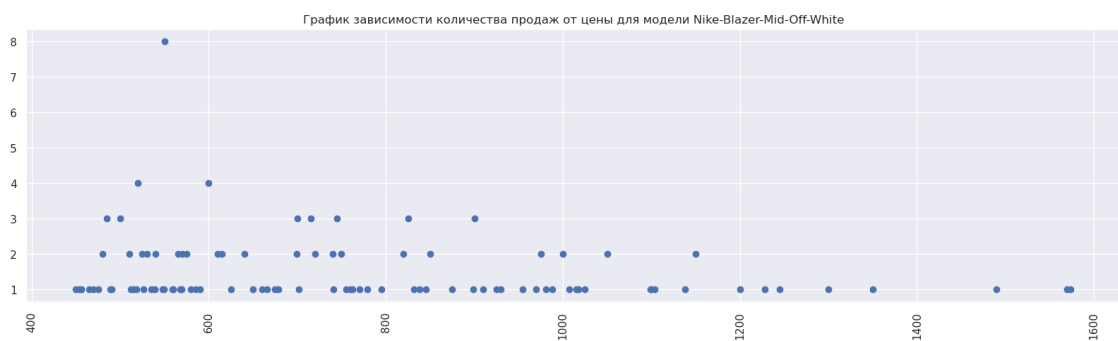
5.0.2

Off-White	- Nike-Air-Max-90-Off-White
Off-White	- Nike-Air-Presto-Off-White
Off-White	- Nike-Air-VaporMax-Off-
White	
Off-White	- Nike-Blazer-Mid-Off-White
Off-White	- Air-Jordan-1-Retro-High-Off-
White-Chicago	
Off-White	- Nike-Air-Force-1-Low-Off-
White	
Off-White	- Nike-Air-Force-1-Low-
Virgil-Abloh-Off-White-AF100	
Off-White	- Nike-Air-Max-97-Off-White
Off-White	- Nike-Zoom-Fly-Off-White
Off-White	- Nike-React-
Hyperdunk-2017-Flyknit-Off-White	
Off-White	- Nike-Air-VaporMax-Off-
White-2018	
Off-White	- Air-Jordan-1-Retro-High-
Off-White-White	
Off-White	- Nike-Air-VaporMax-Off-White-
Black	
Off-White	- Nike-Air-Presto-Off-White-
Black-2018	

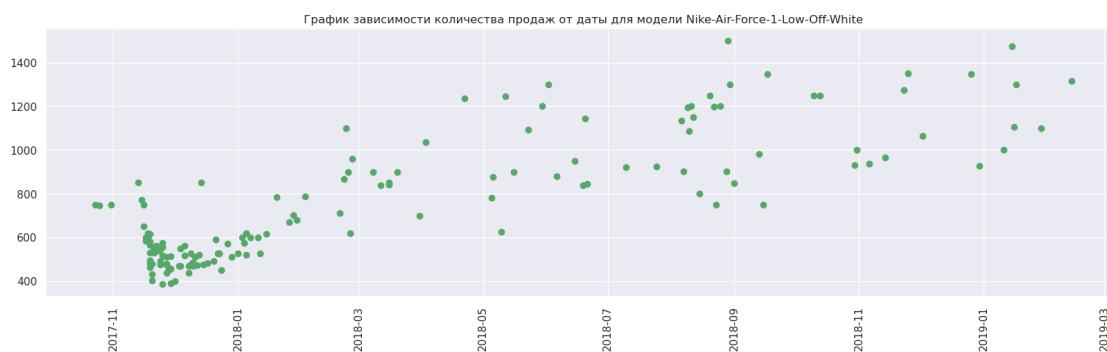
Off-White	-	Air-Jordan-1-Retro-High-
Off-White-University-Blue	-	Nike-Zoom-Fly-Mercurial-
Off-White	-	Nike-Zoom-Fly-Mercurial-
Off-White-Black	-	Nike-Zoom-Fly-Mercurial-
Off-White	-	Nike-Air-Presto-Off-White-
Off-White-Total-Orange	-	Nike-Air-Presto-Off-White-
White-2018	-	Nike-Air-Presto-Off-White-
Off-White	-	Nike-Air-Max-97-Off-White-
Elemental-Rose-Queen	-	Nike-Air-Max-97-Off-White-
Off-White	-	Nike-Blazer-Mid-Off-White-
All-Hallows-Eve	-	Nike-Blazer-Mid-Off-White-
Off-White	-	Nike-Blazer-Mid-Off-White-
Grim-Reaper	-	Nike-Blazer-Mid-Off-White-
Off-White	-	Nike-Blazer-Mid-Off-White-
Wolf-Grey	-	Nike-Blazer-Mid-Off-White-
Off-White	-	Nike-Air-Max-97-Off-White-
Black	-	Nike-Air-Max-97-Off-White-
Off-White	-	Nike-Air-Max-97-Off-White-
Menta	-	Nike-Air-Max-97-Off-White-
Off-White	-	Nike-Zoom-Fly-Off-White-
Black-Silver	-	Nike-Zoom-Fly-Off-White-
Off-White	-	Nike-Zoom-Fly-Off-White-
Pink	-	Nike-Zoom-Fly-Off-White-
Off-White	-	Nike-Air-Force-1-Low-Off-
White-Volt	-	Nike-Air-Force-1-Low-Off-
Off-White	-	Nike-Air-Force-1-Low-Off-
White-Black-White	-	Nike-Air-Force-1-Low-Off-
Off-White	-	Nike-Air-Force-1-Low-Off-
Black	-	Nike-Air-Force-1-Low-Off-
Off-White	-	Nike-Air-Force-1-Low-Off-
Desert-Ore	-	Nike-Air-Force-1-Low-Off-

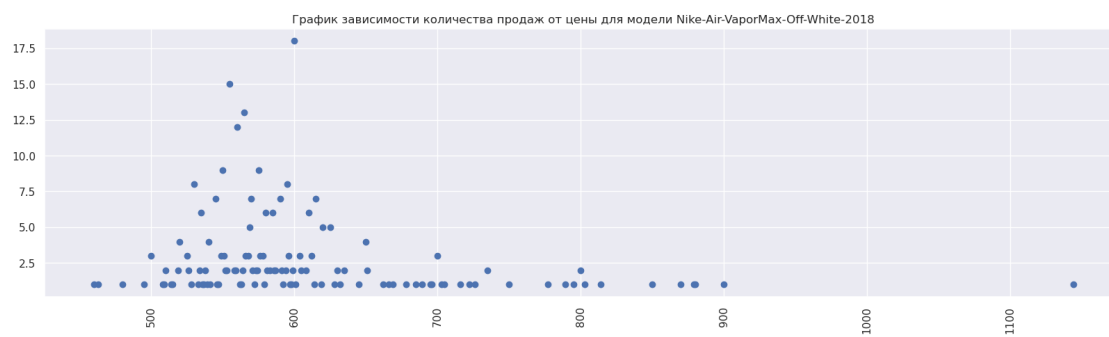
Off-White

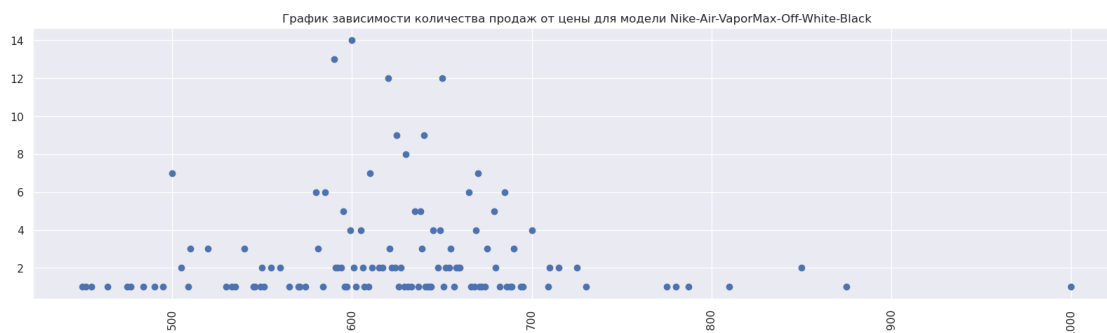










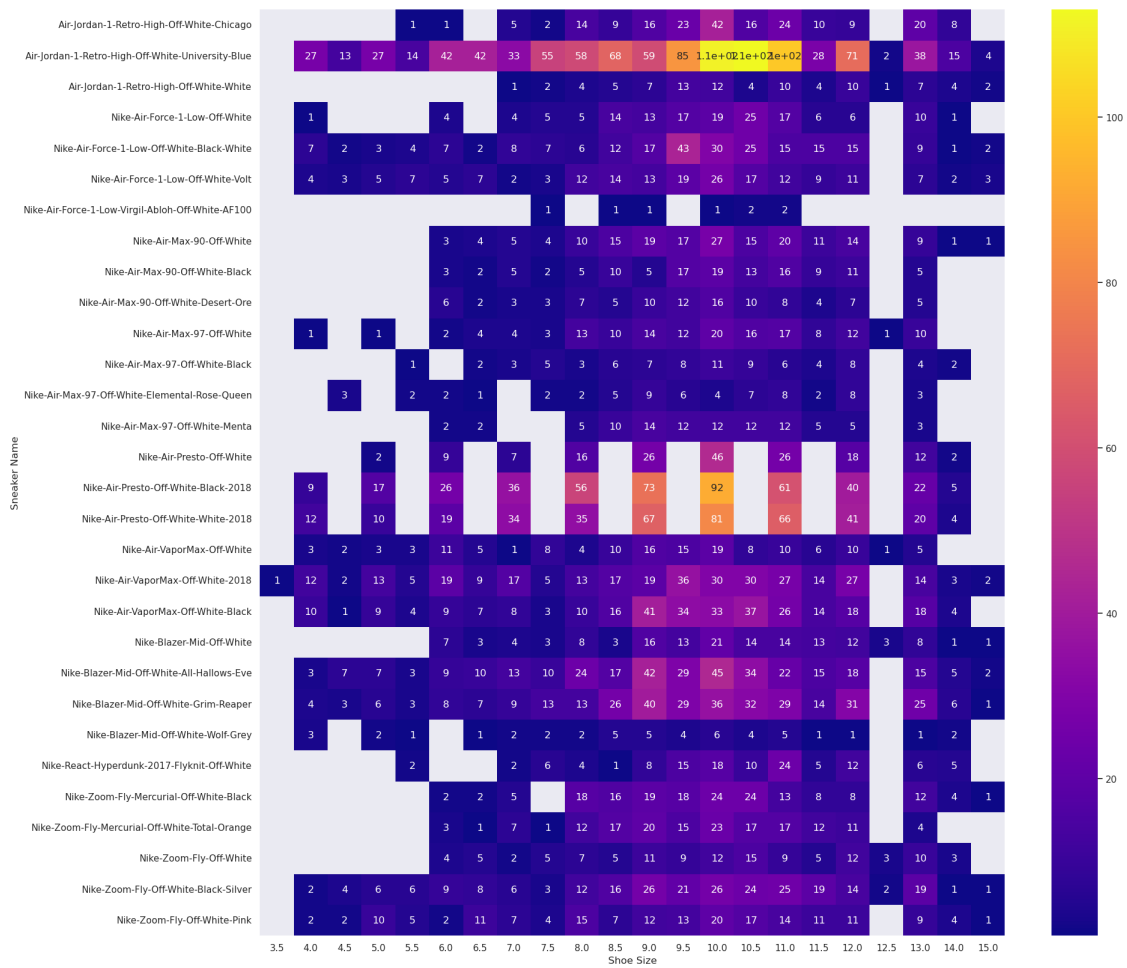




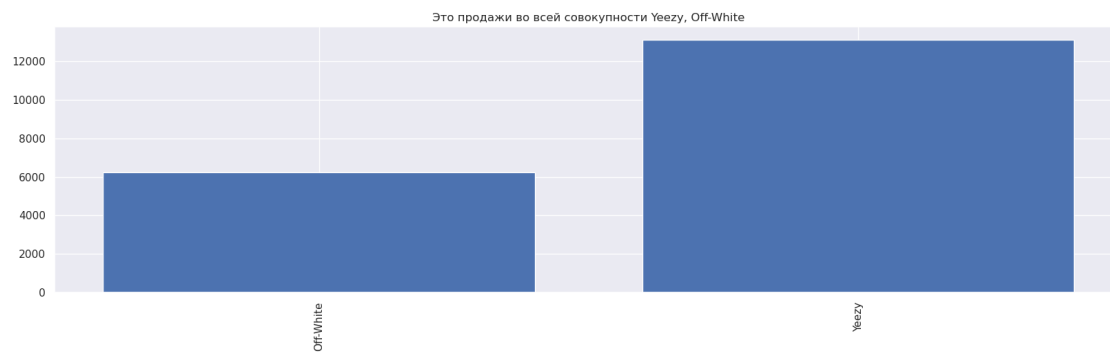
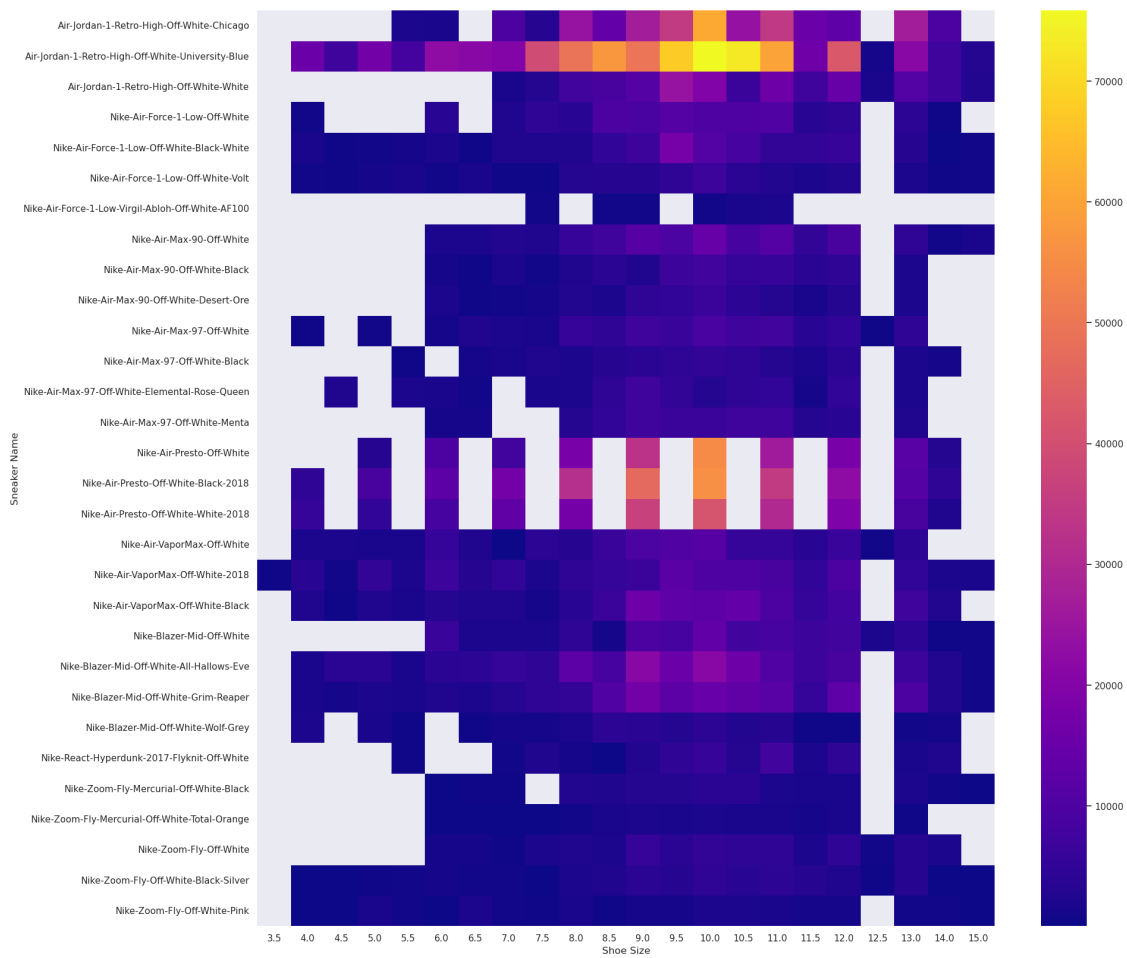
Off-White



Off-White



Off-White



[]: