

## ייבוא חבילות

```
1 import numpy as np
2 import pandas as pd
3 import seaborn as sns
4 import matplotlib.pyplot as plt
5 import datetime
6 from functools import reduce
7 pd.options.display.float_format = '{:,.2f}'.format
```

: [1] In

## בניית פונקציות שימושיות

```
1 def FGname_var(xx):
2     for objname, oid in globals().items():
3         if oid is xx and not objname.startswith("_") :
4             return objname
5     return
```

: [2] In

```
1 def byhourdf(df):
2     FGname_var(df)
3     df2=df.copy()[ (df['Timestamp'].dt.month == 10) & (df['Timestamp'].dt.day==25)]
4     df2['Timestamp_RT']=df2['Timestamp'].dt.floor('H')
5     return df2.groupby(df2['Timestamp_RT']).mean()
```

: [3] In

## ייבוא הדאטה (טבלאות עם נתונים שעתיים שנלקחו מאתר blockchain.com)

```
1 # The total USD value of bitcoin in circulation.
2 market_cap = pd.read_csv('daily/market-cap.csv', parse_dates=['Timestamp'])
```

: [4] In

```
1 transactions_per_second= pd.read_csv('daily/transactions-per-second.csv', parse_dates=['Timestamp'])
```

: [5] In

```
1 total_bitcoins = pd.read_csv('daily/total-bitcoins.csv', parse_dates=['Timestamp'])
```

: [6] In

```
1 my_wallet_n_users = pd.read_csv('daily/my-wallet-n-users.csv', parse_dates=['Timestamp'])
```

: [7] In

```
1 dfs=[market_cap,transactions_per_second,total_bitcoins,my_wallet_n_users]
```

: [8] In

## ניתוח ראשוני של הדאטה

```

1 for df in dfs:
2     print(FGname_var(df))
3     print(df.dtypes)
4     print('\n')
5     print('Nan values')
6     print(df.isnull().sum().sum())
7     print(df.head(4))
8     print('\n')

```

```

market_cap
[Timestamp      datetime64[ns
market-cap      float64
dtype: object

```

```

Nan values
0
Timestamp      market-cap
197,747,000,000.00 01:05:00 2020-09-26 0
198,025,000,000.00 01:26:00 2020-09-26 1
198,025,000,000.00 01:29:00 2020-09-26 2
197,442,000,000.00 01:59:00 2020-09-26 3

```

```

transactions_per_second
[Timestamp      datetime64[ns
transactions-per-second  float64
dtype: object

```

```

Nan values
0
Timestamp  transactions-per-second
3.77      15:30:00 2020-09-26 0
4.02      16:00:00 2020-09-26 1
4.00      16:30:00 2020-09-26 2
4.20      17:00:00 2020-09-26 3

```

```

total_bitcoins
[Timestamp      datetime64[ns
total-bitcoins  float64
dtype: object

```

```

Nan values
0
Timestamp  total-bitcoins
18,500,968.75 01:05:00 2020-09-27 0
18,500,987.50 02:29:00 2020-09-27 1
18,501,006.25 03:35:00 2020-09-27 2
18,501,025.00 03:53:00 2020-09-27 3

```

```

my_wallet_n_users
[Timestamp      datetime64[ns
my-wallet-n-users  int64
dtype: object

```

```

Nan values
0
Timestamp  my-wallet-n-users
54079775   01:54:00 2020-09-26 0
54079775   02:54:00 2020-09-26 1
54079775   03:54:00 2020-09-26 2
54079775   04:54:00 2020-09-26 3

```

## קיבוץ נתונים לפי שעה

לחלק מהשדות יש יותר משורה אחת פר שעה, לכן בחרתי להשתמש בממוצע פר שעה.

```

1 dfs_by_h=[]
2 for df in dfs:
3     vars()[FGname_var(df) + '_by_hour']=byhourdf(df)
4     dfs_by_h.append( vars()[FGname_var(df) + '_by_hour'])
5     print(FGname_var(vars()[FGname_var(df) + '_by_hour']))
6     print(vars()[FGname_var(df) + '_by_hour'].tail())
7     print('\n')

```

: [10] In

market\_cap\_by\_hour  
market-cap

	Timestamp_RT
240,740,000,000.00	18:00:00 2020-10-25
241,245,000,000.00	19:00:00 2020-10-25
241,741,000,000.00	20:00:00 2020-10-25
241,265,000,000.00	22:00:00 2020-10-25
241,478,666,666.67	23:00:00 2020-10-25

transactions\_per\_second\_by\_hour  
transactions-per-second

	Timestamp_RT
3.33	19:00:00 2020-10-25
3.12	20:00:00 2020-10-25
2.83	21:00:00 2020-10-25
2.81	22:00:00 2020-10-25
2.42	23:00:00 2020-10-25

total\_bitcoins\_by\_hour  
total-bitcoins

	Timestamp_RT
18,526,690.62	19:00:00 2020-10-25
18,526,718.75	20:00:00 2020-10-25
18,526,737.50	21:00:00 2020-10-25
18,526,765.62	22:00:00 2020-10-25
18,526,803.12	23:00:00 2020-10-25

my\_wallet\_n\_users\_by\_hour  
my-wallet-n-users

	Timestamp_RT
55400672	19:00:00 2020-10-25
55400672	20:00:00 2020-10-25
55400672	21:00:00 2020-10-25
55400672	22:00:00 2020-10-25
55413053	23:00:00 2020-10-25

## חיבור טבלאות ובדיקת נתונים חסרים.

```

1 df_merged = reduce(lambda left,right: pd.merge(left,right,on=['Timestamp_RT'],
2                                           how='outer'), dfs_by_h)

```

: [11] In

```

1 df_merged.sort_values(by='Timestamp_RT', inplace=True)
2 df_merged.isnull().sum().sum()

```

: [12] In

4

Out[12]:

```

1 df_merged[df_merged.isna().any(axis=1)]

```

: [13] In

Out[13]:

my-wallet-n-users	total-bitcoins	transactions-per-second	market-cap	Timestamp_RT
55377364	18,526,268.75	2.11	nan	05:00:00 2020-10-25
55386127	nan	3.02	240,751,000,000.00	12:00:00 2020-10-25
55400672	18,526,606.25	3.72	nan	16:00:00 2020-10-25
55400672	18,526,737.50	2.83	nan	21:00:00 2020-10-25

## תחקור השדות עם הערכים החסרים על מנת לבחור את הדרך הטובה ביותר למלא אותם

יש כמה שיטות, אציין רק שתיים: האחת, בהסתכלות על הגרפים למטה נראה שגרסיה ליניארית תביא תוצאה טובה לגבי שדה

"total-bitcoins"

אבל תוצאה לא טובה לשדה

"market-cap"

שיטה נוספת היא אינטרפולציה ליניארית אשר תביא לתוצאה טובה והגיוינית לשני השדות עם ערכים חסרים

כאשר המשתנים רציפים עם מגמה ברורה לאורך ציר הזמן, כל פיתרון אחר כגון ממוצע או חציון יכולים לשבש לגמרי את השונוות ולגרום לכך שהשדה יהפוך ללא רלוונטי סטטיסטית

```
1 df_merged.reset_index(inplace=True)
```

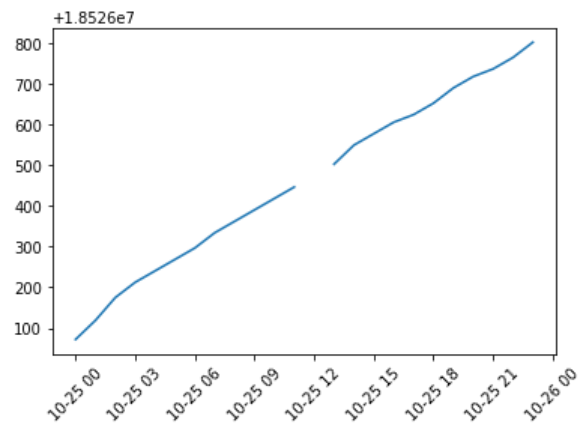
: [14] In

```
1 plt.plot(df_merged['Timestamp_RT'], df_merged['total-bitcoins'])
2 plt.xticks(rotation=45)
```

: [15] In

```
, array([737723.    , 737723.125, 737723.25 , 737723.375, 737723.5)
, ([ 737724 , 737723.875 , 737723.75 , 737723.625
(<a list of 9 Text xticklabel objects>
```

Out[15]:

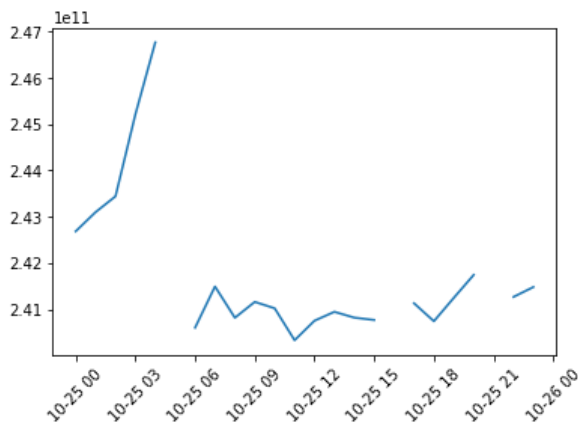


```
1 plt.plot(df_merged['Timestamp_RT'], df_merged['market-cap'])
2 plt.xticks(rotation=45)
```

: [16] In

```
, array([737723.    , 737723.125, 737723.25 , 737723.375, 737723.5)
, ([ 737724 , 737723.875 , 737723.75 , 737723.625
(<a list of 9 Text xticklabel objects>
```

Out[16]:



```
1 df_merged["hour"] = df_merged["Timestamp_RT"].dt.hour
```

: [17] In

```
1 df_merged["market-cap_mus"] = (df_merged['market-cap'].shift(-1) + df_merged['market-cap'].shift(1)) / 2
2 df_merged["total-bitcoins_mus"] = (df_merged['total-bitcoins'].shift(-1) + df_merged['total-bitcoins'].shift(1))
```

: [18] In

```
1 df_merged.head(10)
```

: [19] In

Out[19]:

total-bitcoins_mus	market-cap_mus	hour	my-wallet-n-users	total-bitcoins	transactions-per-second	market-cap	Timestamp_RT	
nan	nan	0	55364970	18,526,071.88	2.38	242,682,500,000.00	2020-10-25 00:00:00	0
18,526,123.44	243,060,500,000.00	1	55364970	18,526,118.75	2.09	243,097,000,000.00	2020-10-25 01:00:00	1
18,526,165.62	244,152,750,000.00	2	55377364	18,526,175.00	2.03	243,438,500,000.00	2020-10-25 02:00:00	2
18,526,207.81	245,101,750,000.00	3	55377364	18,526,212.50	2.06	245,208,500,000.00	2020-10-25 03:00:00	3
18,526,240.62	nan	4	55377364	18,526,240.62	2.17	246,765,000,000.00	2020-10-25 04:00:00	4
18,526,268.75	243,683,000,000.00	5	55377364	18,526,268.75	2.11	nan	2020-10-25 05:00:00	5
18,526,301.56	nan	6	55377364	18,526,296.88	2.31	240,601,000,000.00	2020-10-25 06:00:00	6
18,526,329.69	240,708,000,000.00	7	55377364	18,526,334.38	2.62	241,490,500,000.00	2020-10-25 07:00:00	7
18,526,362.50	241,324,250,000.00	8	55377364	18,526,362.50	2.78	240,815,000,000.00	2020-10-25 08:00:00	8
18,526,390.62	240,917,000,000.00	9	55386127	18,526,390.62	2.83	241,158,000,000.00	2020-10-25 09:00:00	9

```
1 df_merged["market-cap"]=df_merged["market-cap"].fillna(df_merged["market-cap_mus"])
2 df_merged["total-bitcoins"]=df_merged["total-bitcoins"].fillna(df_merged["total-bitcoins_mus"])
```

: [20] In

```
1 df_merged.drop(['total-bitcoins_mus', 'market-cap_mus'], axis=1, inplace=True)
```

: [21] In

1 df\_merged

: [22] In

Out[22]:

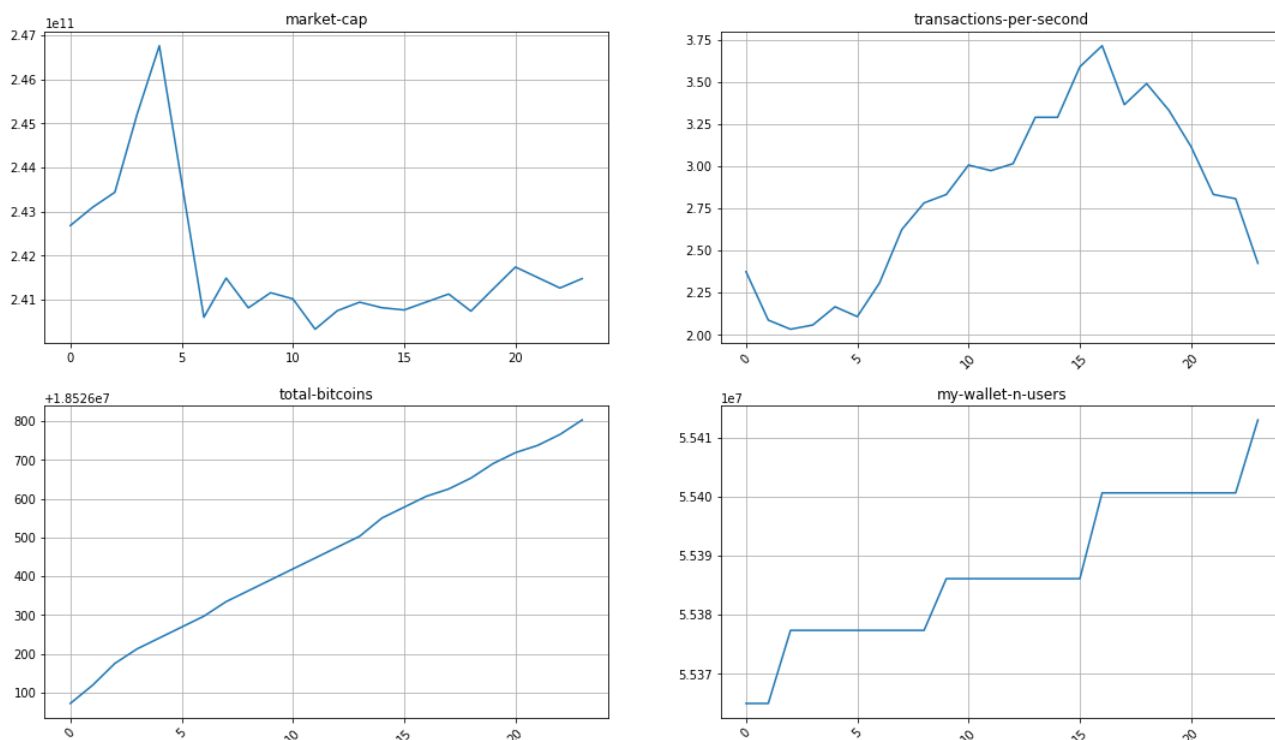
hour	my-wallet-n-users	total-bitcoins	transactions-per-second	market-cap	Timestamp_RT	
0	55364970	18,526,071.88	2.38	242,682,500,000.00	00:00:00 2020-10-25	0
1	55364970	18,526,118.75	2.09	243,097,000,000.00	01:00:00 2020-10-25	1
2	55377364	18,526,175.00	2.03	243,438,500,000.00	02:00:00 2020-10-25	2
3	55377364	18,526,212.50	2.06	245,208,500,000.00	03:00:00 2020-10-25	3
4	55377364	18,526,240.62	2.17	246,765,000,000.00	04:00:00 2020-10-25	4
5	55377364	18,526,268.75	2.11	243,683,000,000.00	05:00:00 2020-10-25	5
6	55377364	18,526,296.88	2.31	240,601,000,000.00	06:00:00 2020-10-25	6
7	55377364	18,526,334.38	2.62	241,490,500,000.00	07:00:00 2020-10-25	7
8	55377364	18,526,362.50	2.78	240,815,000,000.00	08:00:00 2020-10-25	8
9	55386127	18,526,390.62	2.83	241,158,000,000.00	09:00:00 2020-10-25	9
10	55386127	18,526,418.75	3.01	241,019,000,000.00	10:00:00 2020-10-25	10
11	55386127	18,526,446.88	2.98	240,329,500,000.00	11:00:00 2020-10-25	11
12	55386127	18,526,475.00	3.02	240,751,000,000.00	12:00:00 2020-10-25	12
13	55386127	18,526,503.12	3.29	240,943,333,333.33	13:00:00 2020-10-25	13
14	55386127	18,526,550.00	3.29	240,817,500,000.00	14:00:00 2020-10-25	14
15	55386127	18,526,578.12	3.59	240,767,000,000.00	15:00:00 2020-10-25	15
16	55400672	18,526,606.25	3.72	240,947,750,000.00	16:00:00 2020-10-25	16
17	55400672	18,526,625.00	3.37	241,128,500,000.00	17:00:00 2020-10-25	17
18	55400672	18,526,653.12	3.49	240,740,000,000.00	18:00:00 2020-10-25	18
19	55400672	18,526,690.62	3.33	241,245,000,000.00	19:00:00 2020-10-25	19
20	55400672	18,526,718.75	3.12	241,741,000,000.00	20:00:00 2020-10-25	20
21	55400672	18,526,737.50	2.83	241,503,000,000.00	21:00:00 2020-10-25	21
22	55400672	18,526,765.62	2.81	241,265,000,000.00	22:00:00 2020-10-25	22
23	55413053	18,526,803.12	2.42	241,478,666,666.67	23:00:00 2020-10-25	23

הסתכלות על המשתנים שבטבלה

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3
4
5 x=df_merged['hour']
6
7 fig=plt.figure()
8 fig.set_size_inches(18.5, 10.5)
9
10 plt.subplot(221)
11 plt.plot(x, df_merged['market-cap'])
12 plt.yscale('linear')
13 plt.title('market-cap')
14 plt.grid(True)
15
16 plt.subplot(222)
17 plt.plot(x, df_merged['transactions-per-second'])
18 plt.yscale('linear')
19 plt.title('transactions-per-second')
20 plt.grid(True)
21 plt.xticks(rotation=45)
22
23 plt.subplot(223)
24 plt.plot(x, df_merged['total-bitcoins'])
25 plt.yscale('linear')
26 plt.title('total-bitcoins')
27 plt.grid(True)
28 plt.xticks(rotation=45)
29
30 plt.subplot(224)
31 plt.plot(x, df_merged['my-wallet-n-users'])
32 plt.yscale('linear')
33 plt.title('my-wallet-n-users')
34 plt.grid(True)
35 plt.xticks(rotation=45)
36
37 plt.show()

```



שדה "my-wallet-n-users" בולט מאוד לעין שהוא משתנה כל 7 שעות (דבר שלא מתיישב טוב עם המציאות) וגם שהצמיחה שלו היא ליניארית

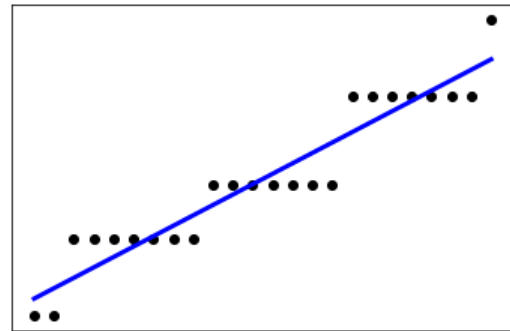
אפשר להריץ רגרסיה ליניארית עבורו ולהשתמש בערכים המנובאים כל עוד הרגרסיה מובהקת

```

1 from sklearn import linear_model
2 from sklearn.metrics import mean_squared_error, r2_score
3 regr = linear_model.LinearRegression()
4 x=np.array(df_merged['hour']).reshape(-1, 1)
5 y=np.array(df_merged['my-wallet-n-users']).reshape(-1, 1)
6 regr.fit(x, y)
7
8 y_pred = regr.predict(x)
9
10 print('Coefficient of determination: %.2f'
11       % r2_score(y, y_pred))
12
13 plt.scatter(x, y, color='black')
14 plt.plot(x, y_pred, color='blue', linewidth=3)
15
16 plt.xticks(())
17 plt.yticks(())
18
19 plt.show()

```

Coefficient of determination: 0.90



R בריבוע יצא מאד גבוה, מה שאומר שמבחינתי הרגרסיה מובהקת ואני יכול להשתמש בערכים המנובאים במקום השדה המקורי ואף לזרוק אותו מהטבלה



```
1 df_merged['est_wallet-n-users']=y_pred
2 df_merged.drop(['my-wallet-n-users'], axis=1, inplace=True)
3 df_merged
```

: [25] In

Out[25]:

est_wallet-n-users	hour	total-bitcoins	transactions-per-second	market-cap	Timestamp_RT	
55,367,797.93	0	18,526,071.88	2.38	242,682,500,000.00	00:00:00 2020-10-25	0
55,369,482.65	1	18,526,118.75	2.09	243,097,000,000.00	01:00:00 2020-10-25	1
55,371,167.38	2	18,526,175.00	2.03	243,438,500,000.00	02:00:00 2020-10-25	2
55,372,852.10	3	18,526,212.50	2.06	245,208,500,000.00	03:00:00 2020-10-25	3
55,374,536.82	4	18,526,240.62	2.17	246,765,000,000.00	04:00:00 2020-10-25	4
55,376,221.55	5	18,526,268.75	2.11	243,683,000,000.00	05:00:00 2020-10-25	5
55,377,906.27	6	18,526,296.88	2.31	240,601,000,000.00	06:00:00 2020-10-25	6
55,379,590.99	7	18,526,334.38	2.62	241,490,500,000.00	07:00:00 2020-10-25	7
55,381,275.72	8	18,526,362.50	2.78	240,815,000,000.00	08:00:00 2020-10-25	8
55,382,960.44	9	18,526,390.62	2.83	241,158,000,000.00	09:00:00 2020-10-25	9
55,384,645.16	10	18,526,418.75	3.01	241,019,000,000.00	10:00:00 2020-10-25	10
55,386,329.89	11	18,526,446.88	2.98	240,329,500,000.00	11:00:00 2020-10-25	11
55,388,014.61	12	18,526,475.00	3.02	240,751,000,000.00	12:00:00 2020-10-25	12
55,389,699.34	13	18,526,503.12	3.29	240,943,333,333.33	13:00:00 2020-10-25	13
55,391,384.06	14	18,526,550.00	3.29	240,817,500,000.00	14:00:00 2020-10-25	14
55,393,068.78	15	18,526,578.12	3.59	240,767,000,000.00	15:00:00 2020-10-25	15
55,394,753.51	16	18,526,606.25	3.72	240,947,750,000.00	16:00:00 2020-10-25	16
55,396,438.23	17	18,526,625.00	3.37	241,128,500,000.00	17:00:00 2020-10-25	17
55,398,122.95	18	18,526,653.12	3.49	240,740,000,000.00	18:00:00 2020-10-25	18
55,399,807.68	19	18,526,690.62	3.33	241,245,000,000.00	19:00:00 2020-10-25	19
55,401,492.40	20	18,526,718.75	3.12	241,741,000,000.00	20:00:00 2020-10-25	20
55,403,177.12	21	18,526,737.50	2.83	241,503,000,000.00	21:00:00 2020-10-25	21
55,404,861.85	22	18,526,765.62	2.81	241,265,000,000.00	22:00:00 2020-10-25	22
55,406,546.57	23	18,526,803.12	2.42	241,478,666,666.67	23:00:00 2020-10-25	23

יצירת שדות חדשים על בסיס השדות הקיימים

אם נחלק את ערכם של כל הביטקוין הקיימים ביחד ("Market\_price" בכמות הביטקוין שקיימים בעולם "total-bitcoins", נקבל את ערך הביטקוין פר שעה

אם נחלק את סך הביטקוין "total-bitcoins" שבשוק בסך הארנקים המנובאים "est\_wallet-n-users", נקבל את ממוצע הכמות הביטקוין פר ארנק בכל שעה ושעה (ערך מצופה)

ההפרש בין כל שורה של משתמשי ארנק ("my-wallet-n-users") לבין השורה הקודמת, נותנת לנו את כמות הארנקים החדשים או כמות הארנקים שנסגרו (אחרי שהחלפנו ערכים מקוריים עם המנובאים מהגרסיה, ברור שהערך מיושר לגמרי)

```
1 df_merged['Market_price']=df_merged['market-cap']/df_merged['total-bitcoins']
2 df_merged['new_bitcoins']=(df_merged['total-bitcoins']-df_merged['total-bitcoins'].shift(1)).fillna(0)
3 df_merged['est_bit_per_wallet']=df_merged['total-bitcoins']/df_merged['est_wallet-n-users']
4 df_merged['est_new-users']=(df_merged['est_wallet-n-users']-df_merged['est_wallet-n-users'].shift(1)).fillna
```

: [26] In

יצירת פאנל סופי שעליו אבצע את הניתוח הסופי

```
1 data_to_analysis=df_merged[['hour','market-cap','transactions-per-second','total-bitcoins',\
2                               'new_bitcoins','est_wallet-n-users','Market_price',\
3                               'est_bit_per_wallet','est_new-users']]
```

: [27] In

```
1 data_to_analysis
```

: [28] In

Out[28]:

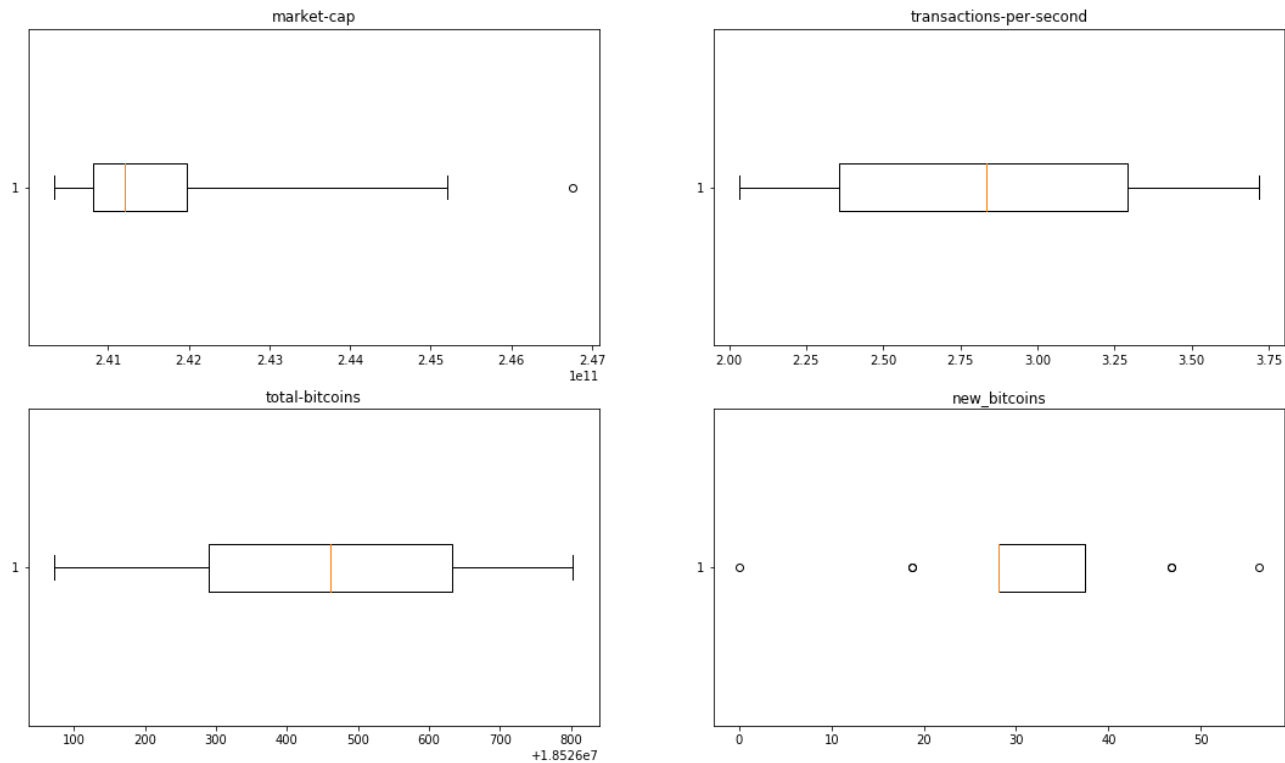
est_new-users	est_bit_per_wallet	Market_price	est_wallet-n-users	new_bitcoins	total-bitcoins	transactions-per-second	market-cap	hour	
0.00	0.33	13,099.51	55,367,797.93	0.00	18,526,071.88	2.38	242,682,500,000.00	0	0
1,684.72	0.33	13,121.85	55,369,482.65	46.88	18,526,118.75	2.09	243,097,000,000.00	1	1
1,684.72	0.33	13,140.25	55,371,167.38	56.25	18,526,175.00	2.03	243,438,500,000.00	2	2
1,684.72	0.33	13,235.76	55,372,852.10	37.50	18,526,212.50	2.06	245,208,500,000.00	3	3
1,684.72	0.33	13,319.76	55,374,536.82	28.12	18,526,240.62	2.17	246,765,000,000.00	4	4
1,684.72	0.33	13,153.38	55,376,221.55	28.12	18,526,268.75	2.11	243,683,000,000.00	5	5
1,684.72	0.33	12,987.00	55,377,906.27	28.12	18,526,296.88	2.31	240,601,000,000.00	6	6
1,684.72	0.33	13,034.99	55,379,590.99	37.50	18,526,334.38	2.62	241,490,500,000.00	7	7
1,684.72	0.33	12,998.50	55,381,275.72	28.12	18,526,362.50	2.78	240,815,000,000.00	8	8
1,684.72	0.33	13,017.00	55,382,960.44	28.12	18,526,390.62	2.83	241,158,000,000.00	9	9
1,684.72	0.33	13,009.48	55,384,645.16	28.12	18,526,418.75	3.01	241,019,000,000.00	10	10
1,684.72	0.33	12,972.24	55,386,329.89	28.12	18,526,446.88	2.98	240,329,500,000.00	11	11
1,684.72	0.33	12,994.97	55,388,014.61	28.12	18,526,475.00	3.02	240,751,000,000.00	12	12
1,684.72	0.33	13,005.33	55,389,699.34	28.12	18,526,503.12	3.29	240,943,333,333.33	13	13
1,684.72	0.33	12,998.51	55,391,384.06	46.88	18,526,550.00	3.29	240,817,500,000.00	14	14
1,684.72	0.33	12,995.76	55,393,068.78	28.12	18,526,578.12	3.59	240,767,000,000.00	15	15
1,684.72	0.33	13,005.50	55,394,753.51	28.12	18,526,606.25	3.72	240,947,750,000.00	16	16
1,684.72	0.33	13,015.24	55,396,438.23	18.75	18,526,625.00	3.37	241,128,500,000.00	17	17
1,684.72	0.33	12,994.25	55,398,122.95	28.12	18,526,653.12	3.49	240,740,000,000.00	18	18
1,684.72	0.33	13,021.48	55,399,807.68	37.50	18,526,690.62	3.33	241,245,000,000.00	19	19
1,684.72	0.33	13,048.24	55,401,492.40	28.12	18,526,718.75	3.12	241,741,000,000.00	20	20
1,684.72	0.33	13,035.38	55,403,177.12	18.75	18,526,737.50	2.83	241,503,000,000.00	21	21
1,684.72	0.33	13,022.51	55,404,861.85	28.12	18,526,765.62	2.81	241,265,000,000.00	22	22
1,684.72	0.33	13,034.02	55,406,546.57	37.50	18,526,803.12	2.42	241,478,666,666.67	23	23

ניתוח סטטיסטיקה תיאורית

```

1 from pylab import *
2
3 fig=plt.figure()
4 fig.set_size_inches(18.5, 10.5)
5
6 subplot(221)
7 boxplot(data_to_analysis['market-cap'],vert=False, whis=4)
8 title('market-cap')
9 subplot(222)
10 boxplot(data_to_analysis['transactions-per-second'],vert=False, whis=0.75)
11 title('transactions-per-second')
12 subplot(223)
13 boxplot(data_to_analysis['total-bitcoins'],vert=False, whis=0.75)
14 title('total-bitcoins')
15 subplot(224)
16 boxplot(data_to_analysis['new_bitcoins'],vert=False, whis=0.75)
17 title('new_bitcoins')
18 show()

```



```

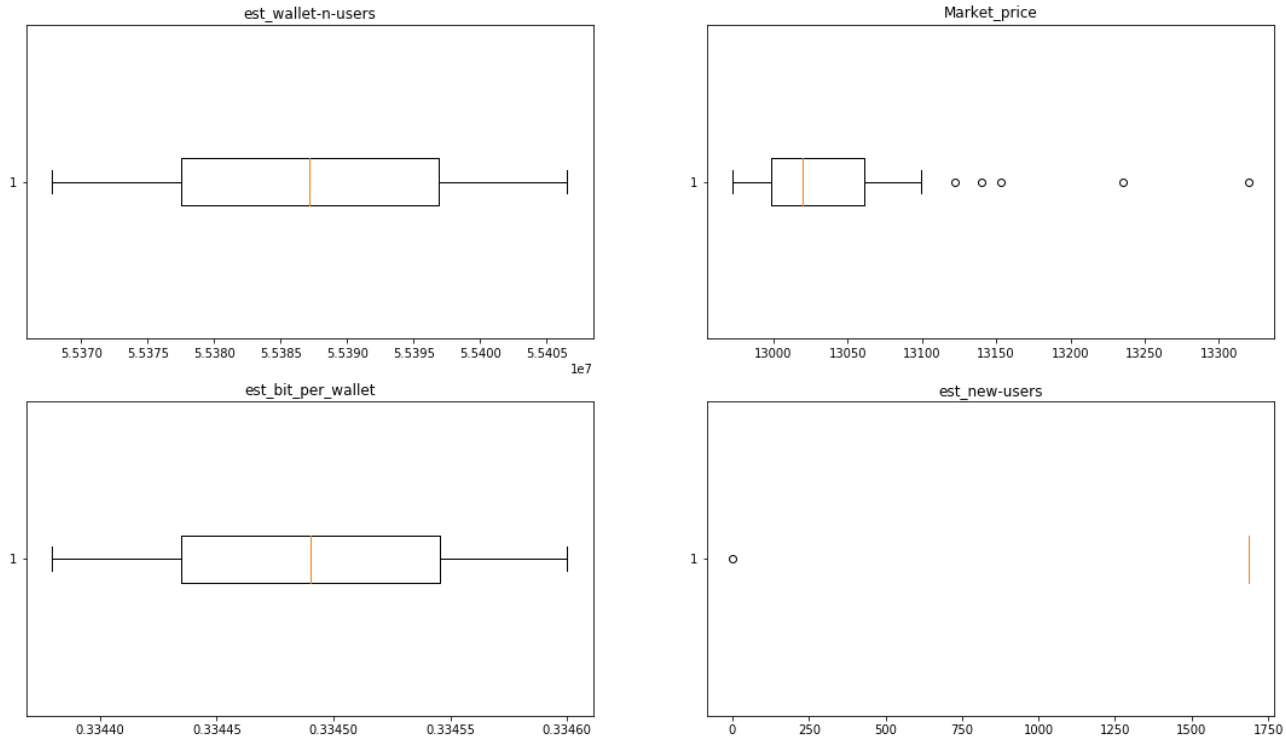
1 fig=plt.figure()
2 fig.set_size_inches(18.5, 10.5)
3
4 subplot(221)
5 boxplot(data_to_analysis['est_wallet-n-users'] ,vert=False, whis=0.75)
6 title('est_wallet-n-users')
7 subplot(222)
8 boxplot(data_to_analysis['Market_price'] ,vert=False, whis=0.75)
9 title('Market_price')
10 subplot(223)
11 boxplot(data_to_analysis['est_bit_per_wallet'] ,vert=False, whis=0.75)
12 title('est_bit_per_wallet')
13 subplot(224)
14 boxplot(data_to_analysis['est_new-users'] ,vert=False, whis=0.75)
15 title('est_new-users')

```

: [30] In

('Text(0.5, 1.0, 'est\_new-users

Out[30]:



```

1 data_to_analysis.iloc[:,1:].describe()

```

: [31] In

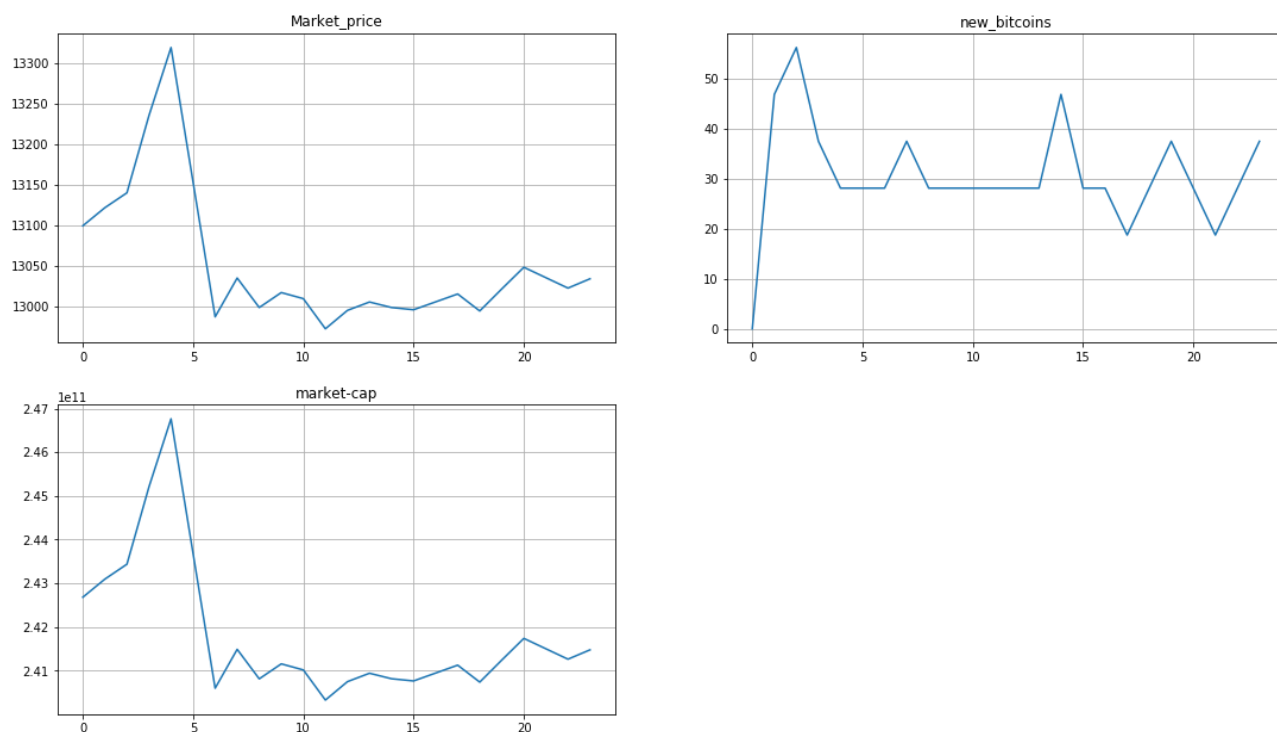
Out[31]:

est_new-users	est_bit_per_wallet	Market_price	est_wallet-n-users	new_bitcoins	total-bitcoins	transactions-per-second	market-cap	
24.00	24.00	24.00	24.00	24.00	24.00	24.00	24.00	count
1,614.53	0.33	13,052.54	55,387,172.25	30.47	18,526,460.16	2.82	241,817,302,083.33	mean
343.89	0.00	85.50	11,912.79	10.80	215.81	0.53	1,582,487,337.83	std
0.00	0.33	12,972.24	55,367,797.93	0.00	18,526,071.88	2.03	240,329,500,000.00	min
1,684.72	0.33	12,998.51	55,377,485.09	28.12	18,526,289.84	2.36	240,816,875,000.00	25%
1,684.72	0.33	13,019.24	55,387,172.25	28.12	18,526,460.94	2.83	241,201,500,000.00	50%
1,684.72	0.33	13,061.06	55,396,859.41	37.50	18,526,632.03	3.29	241,976,375,000.00	75%
1,684.72	0.33	13,319.76	55,406,546.57	56.25	18,526,803.12	3.72	246,765,000,000.00	max

```

1 x=data_to_analysis['hour']
2
3 fig=plt.figure()
4 fig.set_size_inches(18.5, 10.5)
5
6 plt.subplot(221)
7 plt.plot(x, data_to_analysis['Market_price'])
8 plt.yscale('linear')
9 plt.title('Market_price')
10 plt.grid(True)
11
12 plt.subplot(222)
13 plt.plot(x, data_to_analysis['new_bitcoins'])
14 plt.yscale('linear')
15 plt.title('new_bitcoins')
16 plt.grid(True)
17
18 plt.subplot(223)
19 plt.plot(x, data_to_analysis['market-cap'])
20 plt.yscale('linear')
21 plt.title('market-cap')
22 plt.grid(True)

```



### מסקנות מניתוח סטטיסטיקה תיאורית:

אפשר להבין שמחיר הביטקוין לא יציב כל כך עם פיקים לפעמים (volatility). מה שמשתקף גם בערך כספי בדולרים של כל הביטקוין בשוק.

שדה כמות ביטקוינים חדשים מאד לא יציב.

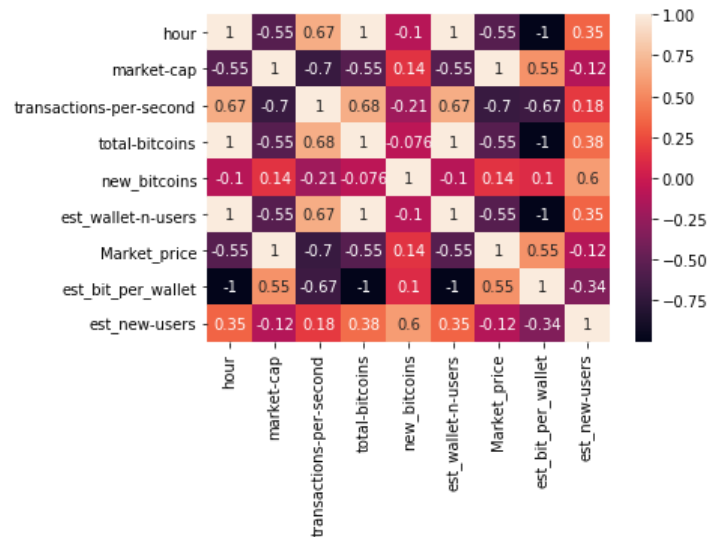
מספר הביטקוין בשוק, מספר הטרנזקציות ומספר הביטקוין פר ארנק יציבים.

שדה של מספר הארנקים החדשים המצופים הוא נבנה מהערכים המנובאים של כמות הארנקים בשוק (שמתעדכן כל כמה שעות כל פעם). מה שמביא למגמה כלפי מעלה ישירה לחלוטין של הארנקים בשוק, דבר זה מאפשר לנו רק להעריך כמה ארנקים מצטרפים פר שעה ולא נותן לנו לדעת בדיוק כמה

נפתחו בפועל. כלומר, מספר הארנקים בשוק עולה בצורה יציבה וישרה ומספר הארנקים החדשים קבוע בכל שעה כי הוא השינוי הממוצע בין שעה לשעה (הדלתא Y הממוצע של קו ישר עולה).

```
1 corrMatrix = data_to_analysis.corr()
2 sns.heatmap(corrMatrix, annot=True)
3 plt.show()
```

In [33]:



## מסקנות ממטריצת קורלציות:

כמות הביטקוין 'total-bitcoins' וכמות הארנקים 'est\_wallet-n-users' בשוק עולים באופן עקבי עם מעבר השעות, וזה מתיישב מאד עם ההיגיון הכללי פיננסי שאומר שבזמני מלחמה ומשבר המשקיעים מעדיפים לברוח מהבורסה והמטבעות, ומתכרזים בנכסים נזילים כמה שיותר ועם כמה שפחות רגולציה (מה יותר טוב ממטבע חופשי מבנק מרכזי כשרוב החנויות המקוונות מקבלות אותו?)

קיים קשר שלילי חזק בין כמות ביטקוין פר ארנק 'est\_bit\_per\_wallet' לבין כמות הארנקים 'est\_wallet-n-users' וגם בין ביטקוין פר ארנק לבין כמות ביטקוין בעולם, וזה מעיד על קצב גידול חזק יותר בכמות המשקיעים מול קצב איטי יותר בכריית המטבע.

הקשר בין מחיר השוק של הביטקוין 'Market\_price' לבין כמות הטרנזקציות לשניה 'transactions-per-second' יכול להסביר את המושגים "Long" ו "Short" אשר מניחים שבתקופות הגאות, השוק עולה לאט ובמצבים של שפל השוק נופל מהר. כאשר מחיר הביטקוין עולה אז כמות הטרנזקציות יורדת, כאשר מחיר השוק יורד כמות הטרנזקציות עולה.

כל הקשרים האחרים בעיני הם לא חזקים (פחות מ-0.7).