به نام خدا

پروژه درس محاسبات آماری

موضوع: آموزش نصب پایتون، آموزش رسم انواع نمودارها در پایتون

محقق: محمد شاهپری

شماره دانشجویی: ۹۶۱۱۳۵۰۰۲۳

سال تحصیلی ۱۴۰۱-۱۴۰۰

دانشگاه سمنان
Semnan University

تاریخچه پایتون

درمورد اینکه تاریخچه **زبان برنامه نویسی پایتون چیست** باید گفت، زبان برنامه نویسی پایتون، در اواخر دهه ۱۹۸۰، توسط گویدو ون راسم «Guido Van Rossum» محقق هلندی ابداع شد.«Van Rossum» در سال ۱۹۸۹ پیاده سازی پایتون را آغاز کرد و آن را جایگزین زبان ABC معرفی کرد. «Van Rossum» در ۱۳ ژانویه ۱۹۵۶ در هلند متولد شد او به دلیل بهره‌مندی از هوش و پشتکار فراوان به موفقیت های بسیاری دست پیدا کرد. «Van Rossum» پایتون را زمانی که در مؤسسه ملی تحقیقات ریاضی و رایانه مشغول به کار بود ابداع کرد. پیش از پایتون زبانی با نام ABC پیاده سازی شد که با مشکلات بسیاری روبرو بود اما برخی ویژگی های این زبان از دید «Van Rossum» خوب و کاربردی بودند. ازین رو او با بهره گیری از تعدادی از ویژگی های ABC، پایتون را پیاده سازی کرد.

مقدمه

- **زبان برنامه نویسی پایتون چیست؟**

به بیان فنی پایتون یک زبان برنامه‌نویسی شی گرا و سطح بالا برای وب و ساخت و توسعه نرم افزار های کاربردی است. پایتون، زبان برنامه‌نویسی نسبتا ساده‌ای محسوب می‌شود که یادگیری آن به دلیل دارا بودن یکتایی که بر قابلیت خوانایی تمرکز و تاکید دارد آسان است. خواندن و ترجمه کدهای نوشته شده به زبان برنامه نویسی پایتون نسبت به دیگر زبان‌ها برای توسعه دهندگان ساده تر محسوب می‌شود. علاوه بر این، زبان برنامه نویسی پایتون از «ماژول‌ها» (modules) و «بسته‌ها» (packages) استفاده می‌کند، بدین معنا که برنامه‌های این زبان قابل طراحی به سبک «ماژولار» (modular) هستند و کدهای نوشته شده در یک پروژه در پروژه‌های گوناگون دیگر نیز قابل استفاده مجدد محسوب می‌شوند. هنگامی که کاربری ماژول یا بسته مورد نیاز خود را توسعه داد، خودش یا دیگر علاقمندان (در صورتی که کد در اختیار عموم قرار بگیرد) می‌توانند آن را برای استفاده در دیگر پروژه‌ها گسترش دهند. «ایمپورت» (Import) و «اکسپورت» (Export) کردن این ماژول‌ها نیز کار آسانی است. در ادامه به بررسی شیوه نصب پایتون و آناکوندا، اهمیت پایتون، رسم نمودارهای آماری مهم و شیوه نصب پکیج در پایتون می‌پردازیم.

اهمیت پایتون

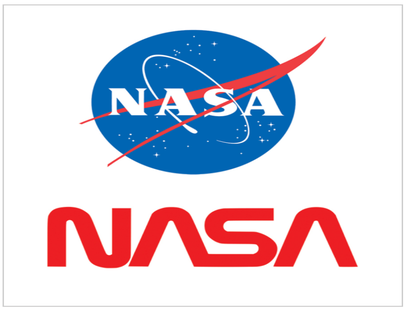
- چرا پایتون در بین تمام زبان های برنامه نویسی محبوب است؟

← غول های فناوری دنیا این زبان را دوست دارند

← مناسب بودن برای مبتدی ها

← پایتون می‌تواند تنها زبان مورد نیاز کاربر باشد

← یعنی پایتون قابلیت‌های زبان های برنامه نویسی گوناگون را دارا است و بنابراین می‌تواند تنها زبانی باشد که یک برنامه‌نویس برای کلیه وظایف لازم در حین توسعه به آن نیاز دارد.



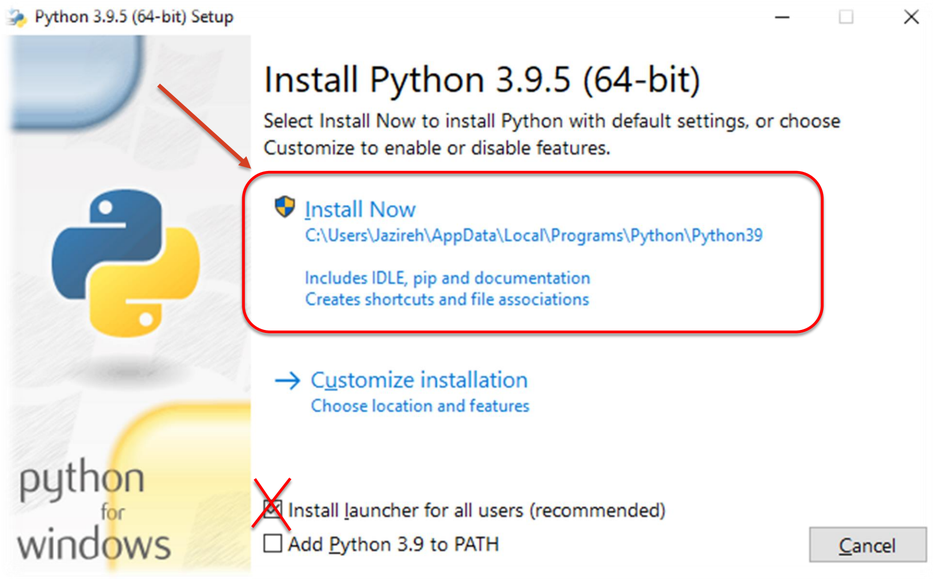
نصب پایتون در ویندوز

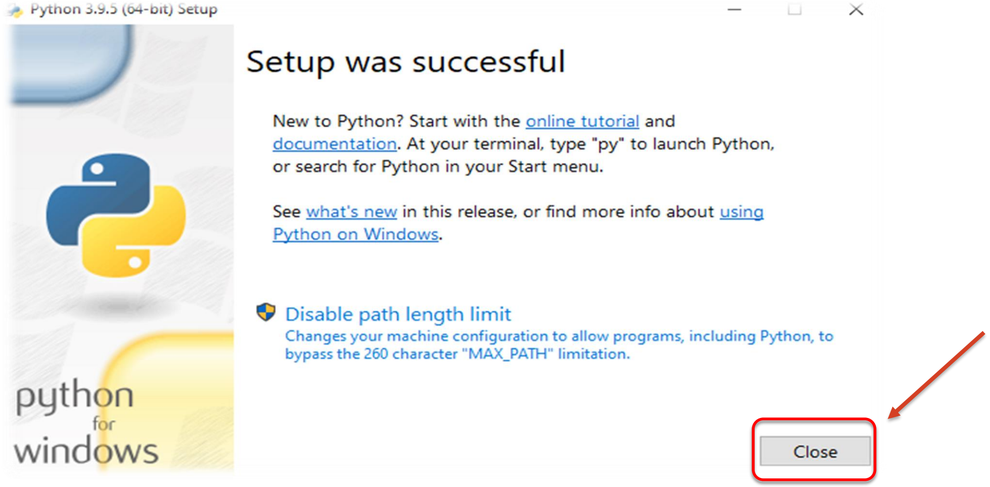
- انتخاب نسخه مناسب برای نصب پایتون در ویندوز

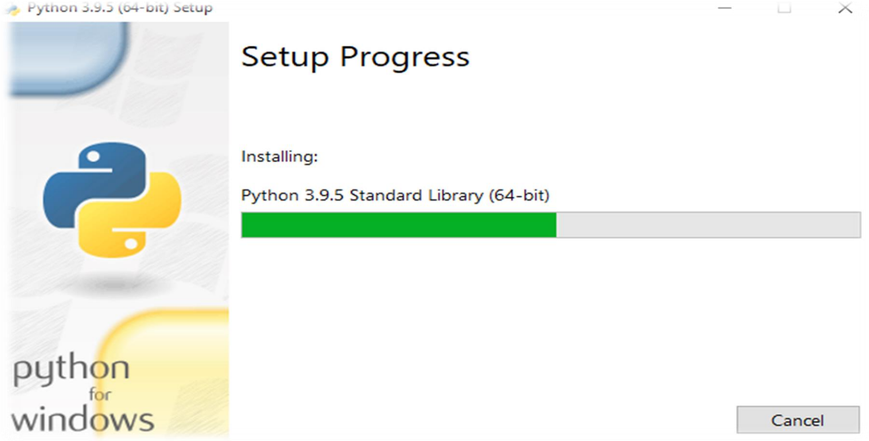
بنیاد نرم‌افزاری پایتون

❖ Python Software Foundation

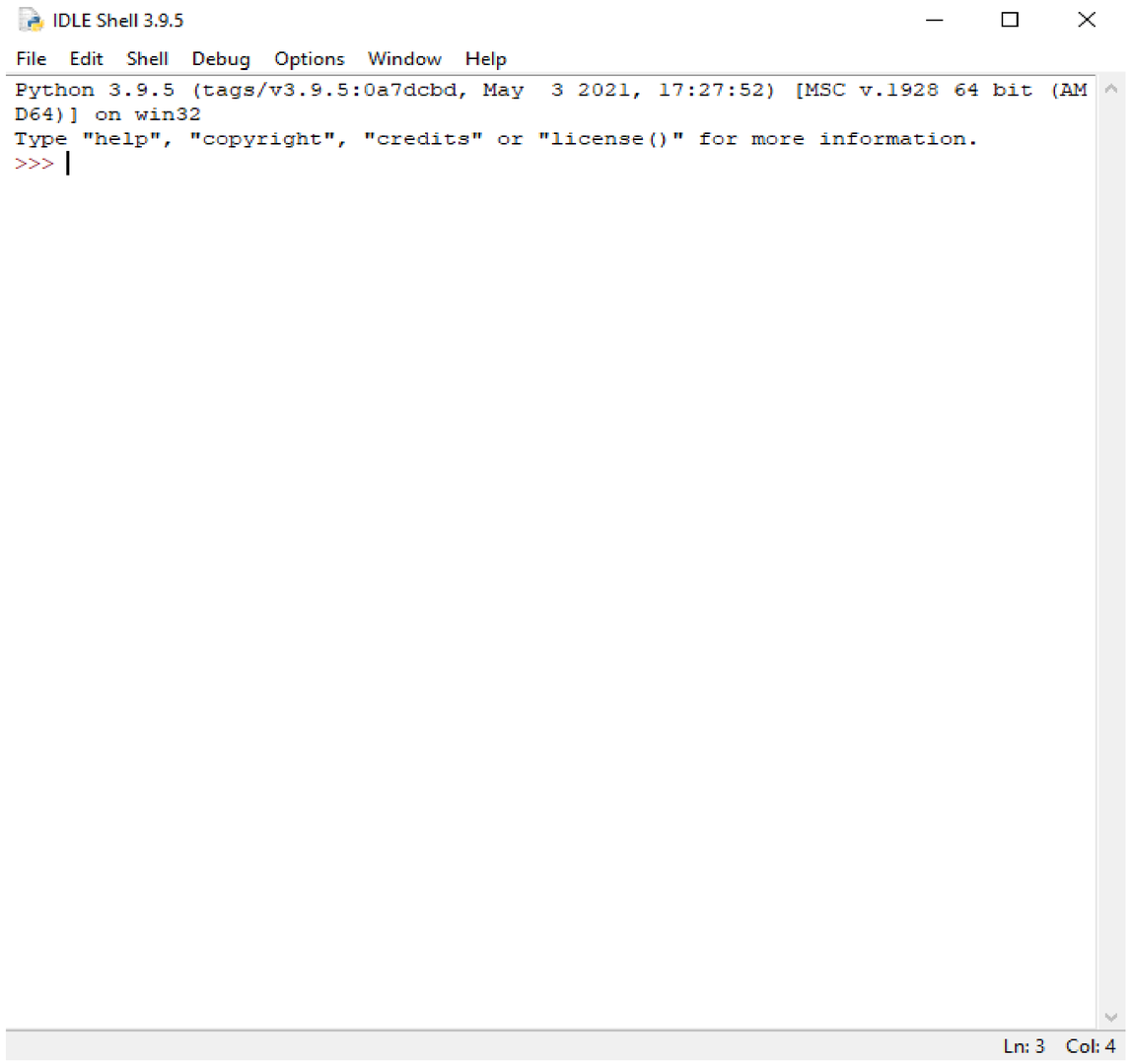
Python 3.9


Python 3.9.5 (64-bit) Setup
Install Python 3.9.5 (64-bit)
Select Install Now to install Python with default settings, or choose
Customize to enable or disable features.
Install Now
C:\Users\Jazireh\AppData\Local\Programs\Python\Python39
Includes IDLE, pip and documentation
Creates shortcuts and file associations
Customize installation
Choose location and features
python
for
windows
Install launcher for all users (recommended)
Add Python 3.9 to PATH
Cancel


Python 3.9.5 (64-bit) Setup
Setup was successful
New to Python? Start with the online tutorial and
documentation. At your terminal, type "py" to launch Python,
or search for Python in your Start menu.
See what's new in this release, or find more info about using
Python on Windows.
Disable path length limit
Changes your machine configuration to allow programs, including Python, to
bypass the 260 character "MAX_PATH" limitation.
python
for
windows
Close


Setup Progress
Installing:
Python 3.9.5 Standard Library (64-bit)
python
for
windows
Cancel

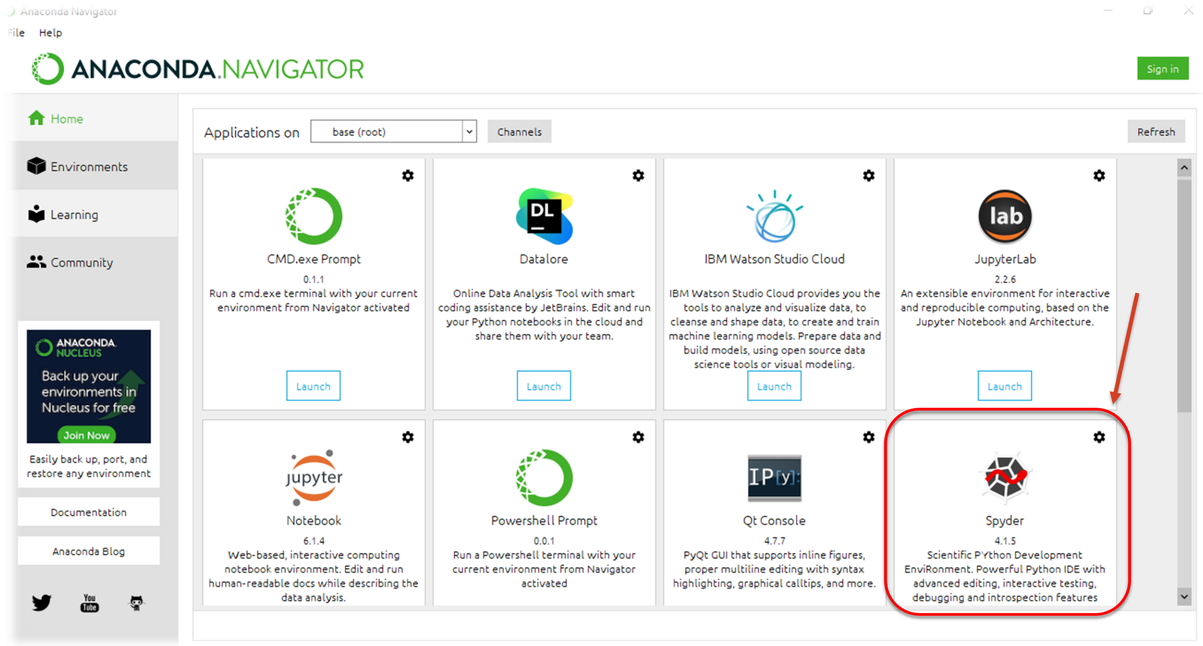
Python 3.9



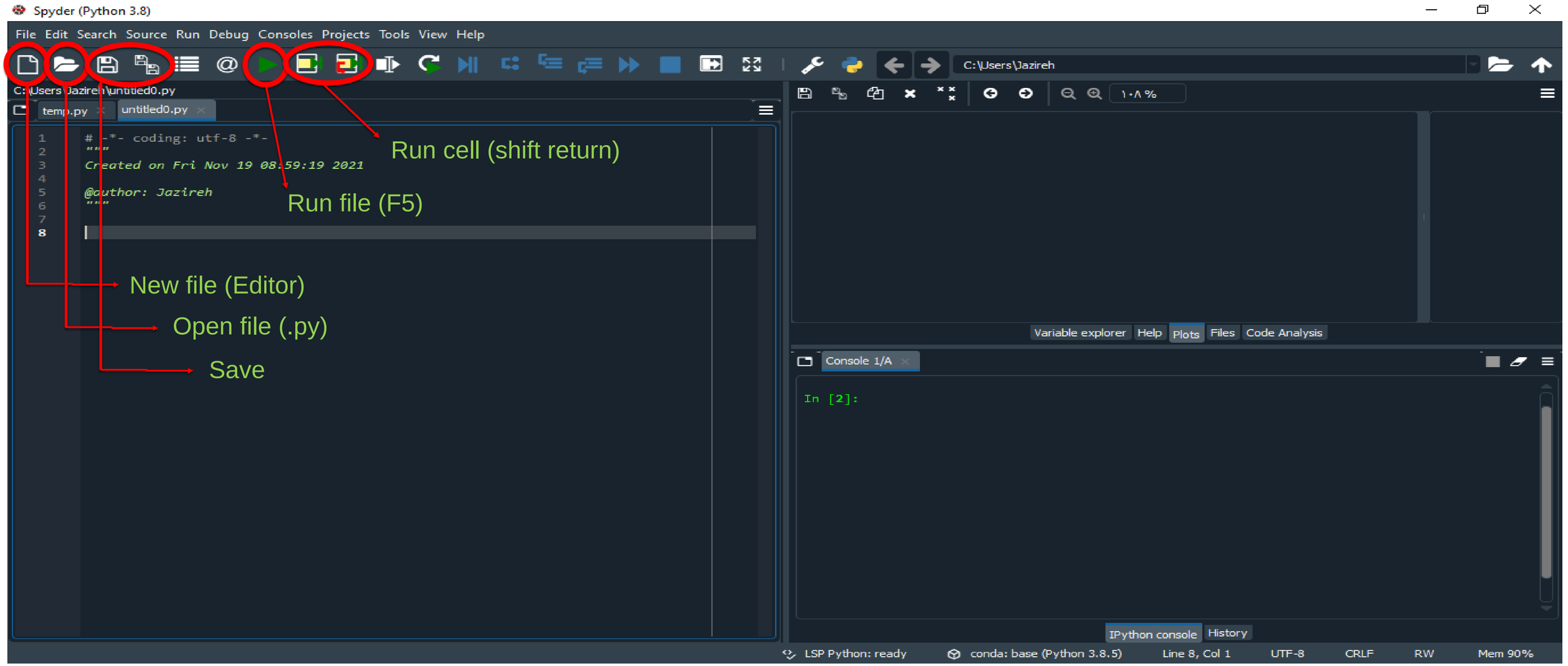
PIP برای پایتون چیست؟

PIP در واقع یک ابزار خط فرمان است که بسته‌های PyPI را با یک دستور ساده و سرراست به نام «PIP» نصب، حذف و یا نصب مجدد می‌کند. اگر از نسخه پایتون ۲.۷ به بالا استفاده کنید PIP بطور پیش‌فرض نصب می‌شود.

Anaconda-spyder



آشنایی با محیط برنامه نویسی اسپایدر



مقایسه زبان برنامه‌نویسی پایتون با زبان C

- برنامه ای بنویسید که یک کلمه یا جمله را دریافت کند و آن را در خروجی به ما نشان دهد.

C programming

```
Int main()
{
Print(“Hello”);
return 0;
}
```

Python programming

```
print(“Hello”)
```

خطاها در نرم‌افزار پایتون

برخی از خطاهایی که ممکن است در پایتون با آن روبرو شویم.

- خطاهای نحوی (غلط املایی)
- خطاهای زمان اجرا (عدم اجرا برنامه)
- خطاهای معنایی یا منطقی (عدم اجرا **صحیح** برنامه، خروجی برنامه دور از انتظار شما است)
- خطا نحو (syntaxError: invalid syntax) (عدم رعایت قواعد زبان برنامه‌نویسی)
- خطا اسمی (NameError: name “name” is not defined) (استفاده از متغیر تعریف نشده)
- خطا نوع (TypeError: object of type 'int' has no len()) (استفاده از یک مقدار نادرست در یک عملیات، تابع و...)

پکیج ها در پایتون

- نصب ۷۰۰ پکیج کاربردی با نصب نرم افزار(Anaconda)
- numpy
- pandas
- scipy
- و...

❖ نمودار ها
❖ توابع (mean,variance,shapiro)
❖ و...

❑ Import __ as __ و from __ import __

رسم انواع نمودارها در پایتون

- شیوه کد نویسی
- نحوه فراخوانی پکیج

```
x=[10,24,13,15,19,10,29,12]
x

import numpy as np
np.mean(x)
np.var(x)
np.sqrt(x)
```

```
In [20]: x=[10,24,13,15,19,10,29,12]
    ...: x
Out[20]: [10, 24, 13, 15, 19, 10, 29, 12]

In [21]: np.mean(x)
Out[21]: 16.5

In [22]: np.var(x)
Out[22]: 42.25

In [23]: np.sqrt(x)
Out[23]:
array([3.16227766, 4.89897949, 3.60555128, 3.87298335, 4.35889894,
       3.16227766, 5.38516481, 3.46410162])
```

نحوه رسم نمودار های

- تشخیصی
- رگرسیونی
- توصیفی
- و....

فراخوانی داده‌ها

```
import pandas as pd
data=pd.read_csv(r"D:\statistical\trm 9\mohasebat amari\project\data.csv")
print(data)
print(data.Height)

X = data['Height'].values
Y = data['lastHeight'].values
```

```
Height  lastHeight
0     126     135.0
1     133     137.0
2     146     158.0
3     143     156.0
4     130     135.0
..    ...       ...
144   131     143.0
145   120     134.0
146   146     158.0
147    92     107.5
148   147     158.0

[149 rows x 2 columns]
```

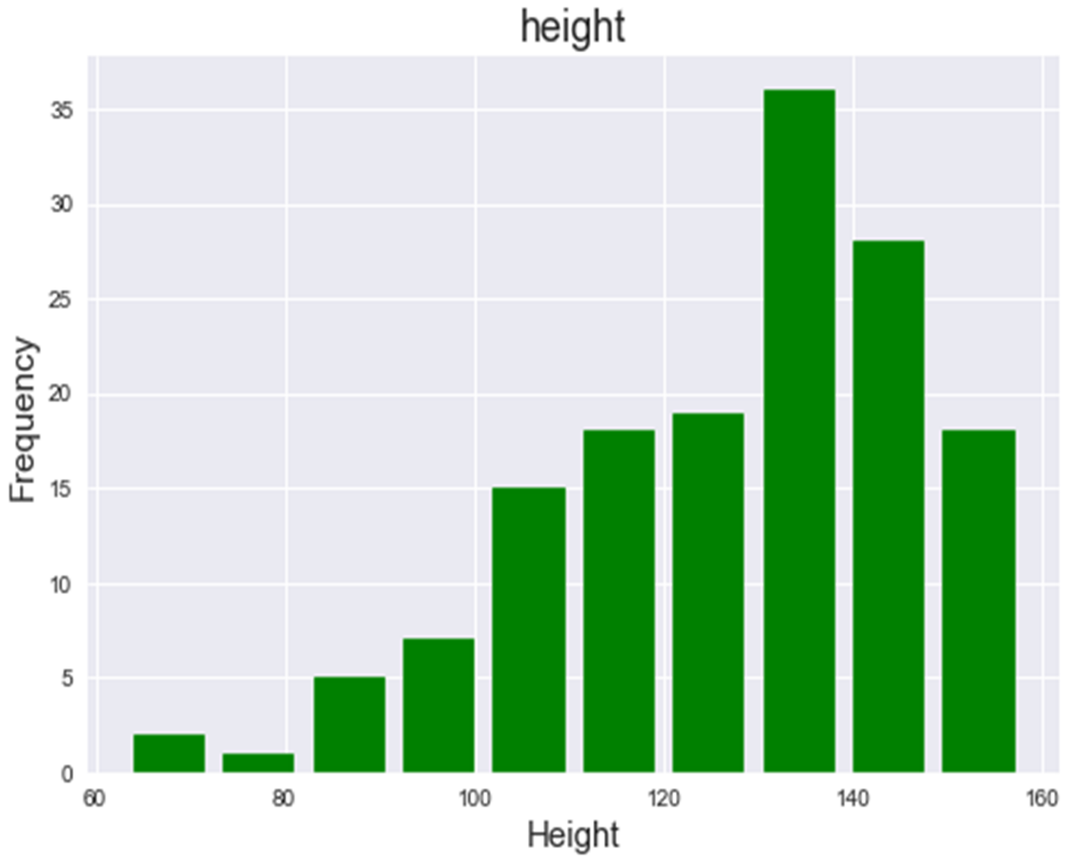
برای شروع کار از داده‌های مربوط به قد کودکان استفاده کردیم. طی یک مطالعه بالینی داروی خاصی را (دارو افزایش قد) بین ۱۵۰ نفر از کودکانی که قد آنها رشد قابل ملاحظه‌ای نکرده است توزیع کردیم. داده‌ها شامل دو متغیر قد قبل از مصرف دارو (Height) و قد بعد از مصرف دارو (lastHeight) هستند.

- فراخوانی داده از نرم‌افزار اکسل(csv file) (read_csv)
- انتخاب نام مناسب برای داده‌ها در پایتون (راحت‌تر شدن کد نویسی)

نمودار هیستوگرام

```
import matplotlib.pyplot as plt
# Histogram #
plt.hist(X,color = 'green', rwidth = 0.8)
plt.xlabel("Height")
plt.ylabel("Frequency")
plt.title("height")
plt.show()
```

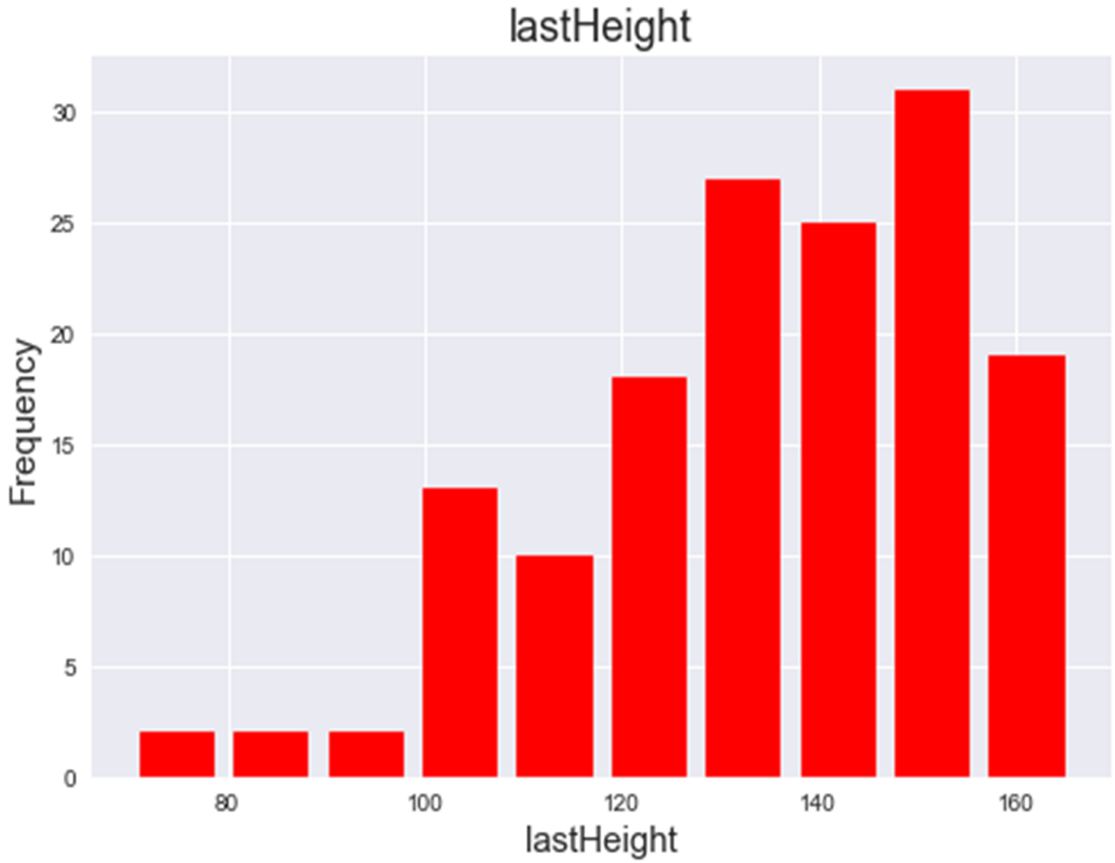
نمودار هیستوگرام مربوط به متغیر X (قد کودکان قبل از مصرف دارو) را مشاهده می‌کنید. بنظر می‌رسد که شکل توزیع داده‌ها کمی چوله به چپ است.



نمودار هیستوگرام

```
# Histogram #
plt.hist(Y,color = 'red', rwidth = 0.8)
plt.xlabel("lastHeight")
plt.ylabel("Frequency")
plt.title("lastHeight")
plt.show()
```

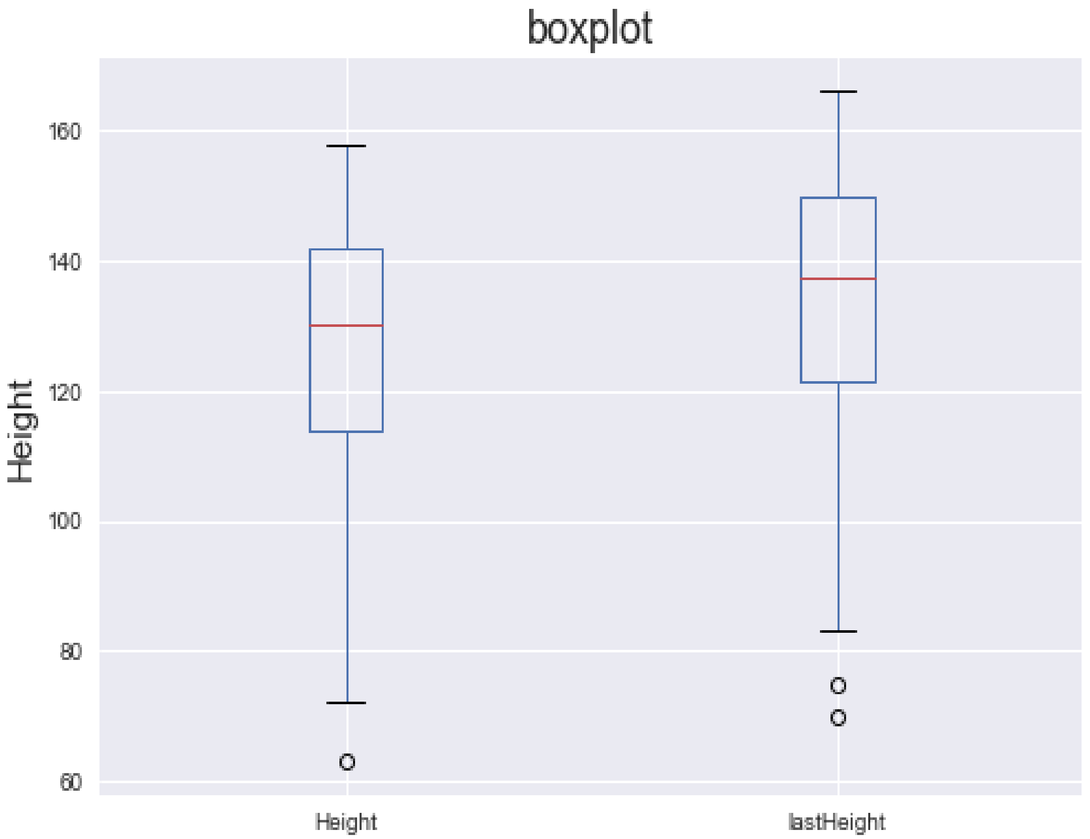
نمودار هیستوگرام مربوط به متغیر Y (قد کودکان بعد از مصرف دارو) را مشاهده می‌کنید. بنظر می‌رسد که شکل توزیع داده‌ها کمی چوله به چپ است.



نمودار جعبه‌ای

```
# boxplot #
data.boxplot(column=["Height","lastHeight"])
plt.ylabel("Height")
plt.title('boxplot')
plt.show()
```

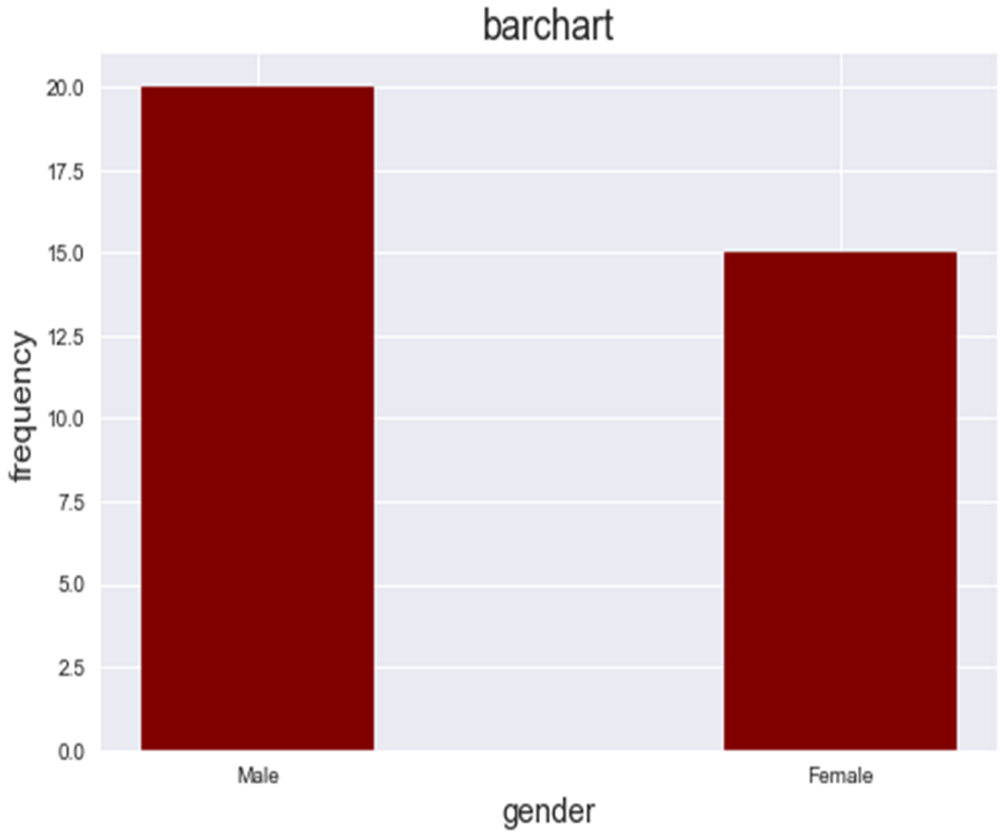
نمودار جعبه‌ای داده‌های مربوط به قد کودکان را مشاهده می‌کنید. بنظر می‌رسد که نمودار جعبه‌ای برای قد کودکان قبل و بعد از مصرف دارو کمی چوله به چپ می‌باشد. همچنین در دو نمودار وجود داده پرت نیز مشاهده می‌شود.



نمودار میله‌ای

```
# barchart #
data1 = {'Male':20, 'Female':15}
gender = list(data1.keys())
values = list(data1.values())
plt.bar(gender, values, color ='maroon', width = 0.4)

plt.xlabel("gender")
plt.ylabel("frequency")
plt.title("barchart")
plt.show()
```

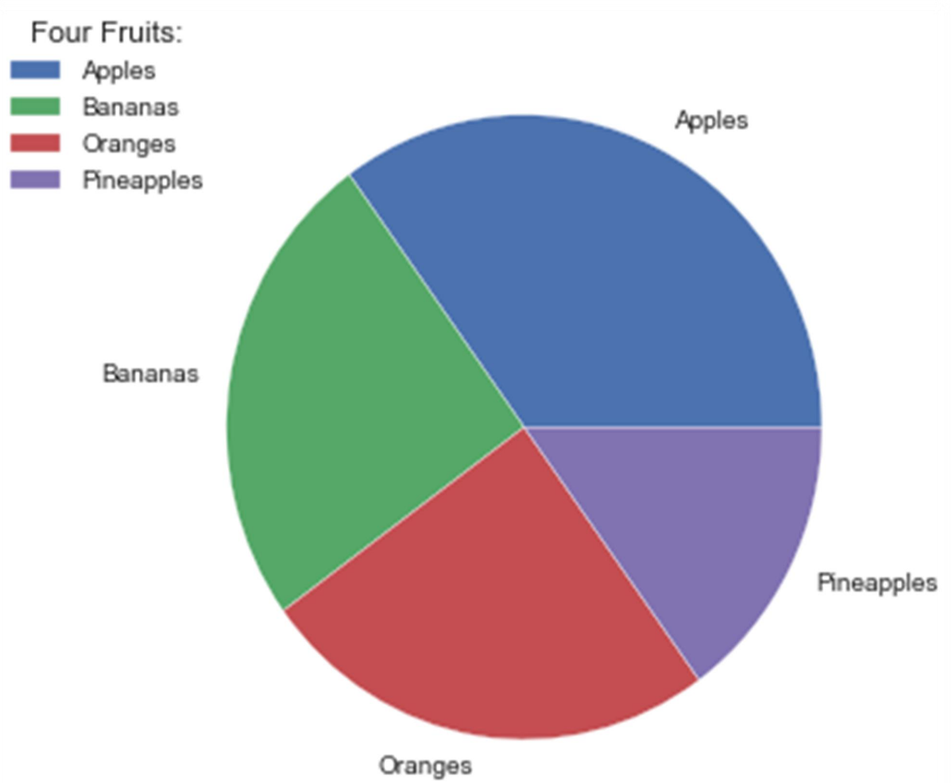
نمودار میله‌ای مربوط به جنسیت را مشاهده می‌کنید. در این نمودار تعداد زنان و مردان مشخص شده است.



نمودار دایره‌ای

```
# piechart #
import matplotlib.pyplot as plt
import numpy as np

z = np.array([35, 25, 25, 15])
mylabels = ["Apples", "Bananas", "Oranges", "Pineapples"]

plt.pie(z, labels = mylabels)
plt.legend(title = "Four Fruits:",
loc='upper right',bbox_to_anchor=(0.1, 1.05))
plt.show()
```

نمودار دایره‌ای مربوط به میوه‌ها را مشاهده می‌کنید. این نمودار سهم هر یک از دسته‌ها (که در اینجا نوع میوه هست) را به ما نشان می‌دهد.



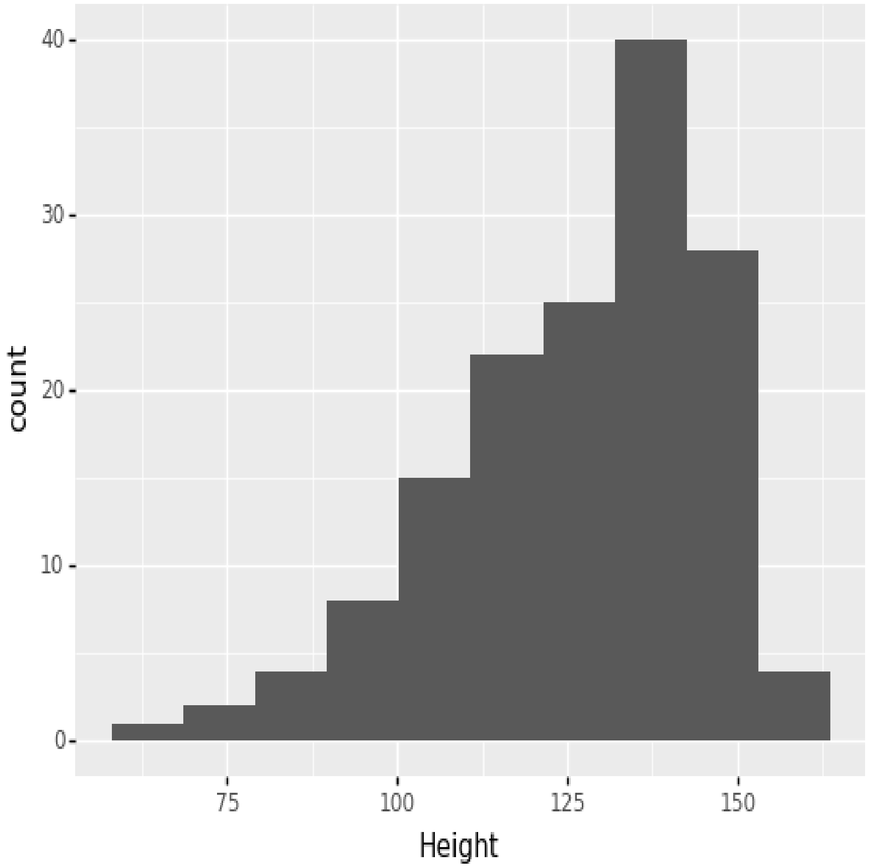
نمودار هیستوگرام

```
# ggplot #
# !pip install pandas plotnine
```

شیوه نصب پکیج

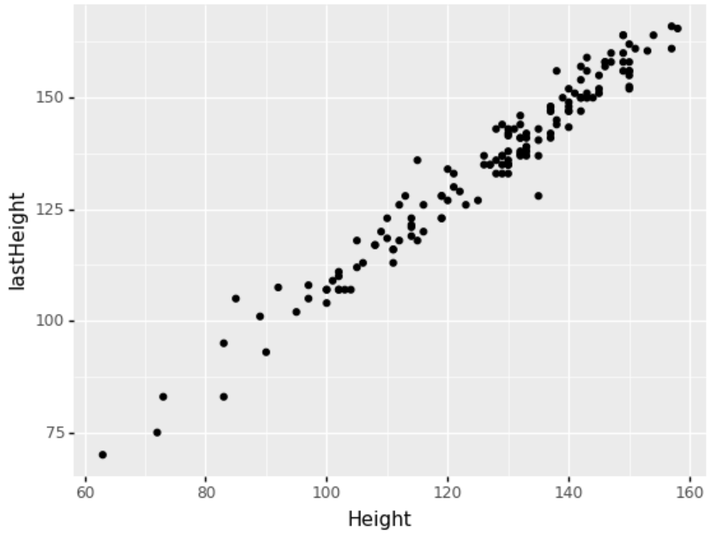
```
from plotnine import ggplot, aes
from plotnine import geom_histogram
from plotnine import geom_boxplot, geom_point

# Historam #
ggplot(data) + aes(x="Height") + geom_histogram(bins=10)
```

در این قسمت با استفاده از تابع ggplot نمودارهای مختلفی را رسم کردیم. با استفاده از تابع ggplot بطور واضح‌تر و دقیق‌تر می‌توان نمودارهای مختلف را مورد بحث و بررسی قرار داد. مشاهده می‌شود که شکل توزیع قد کودکان قبل از مصرف دارو چوله به چپ است.

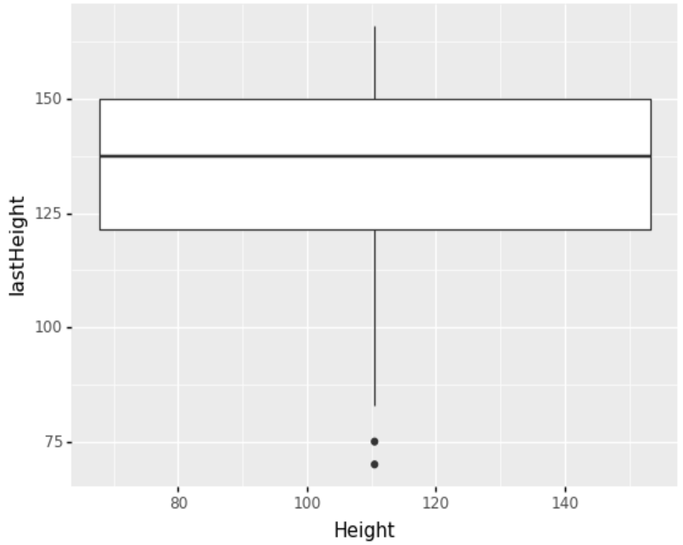


نمودار هیستوگرام، جعبه‌ای و پراکنش

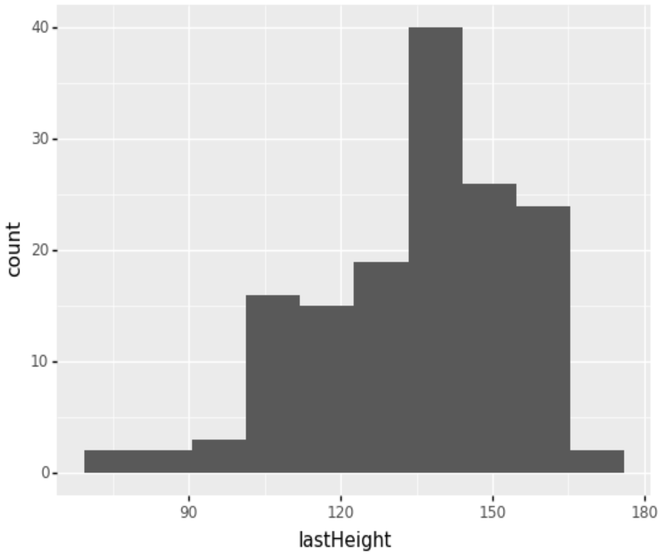
ggpoint



ggboxplot



gghistogramt

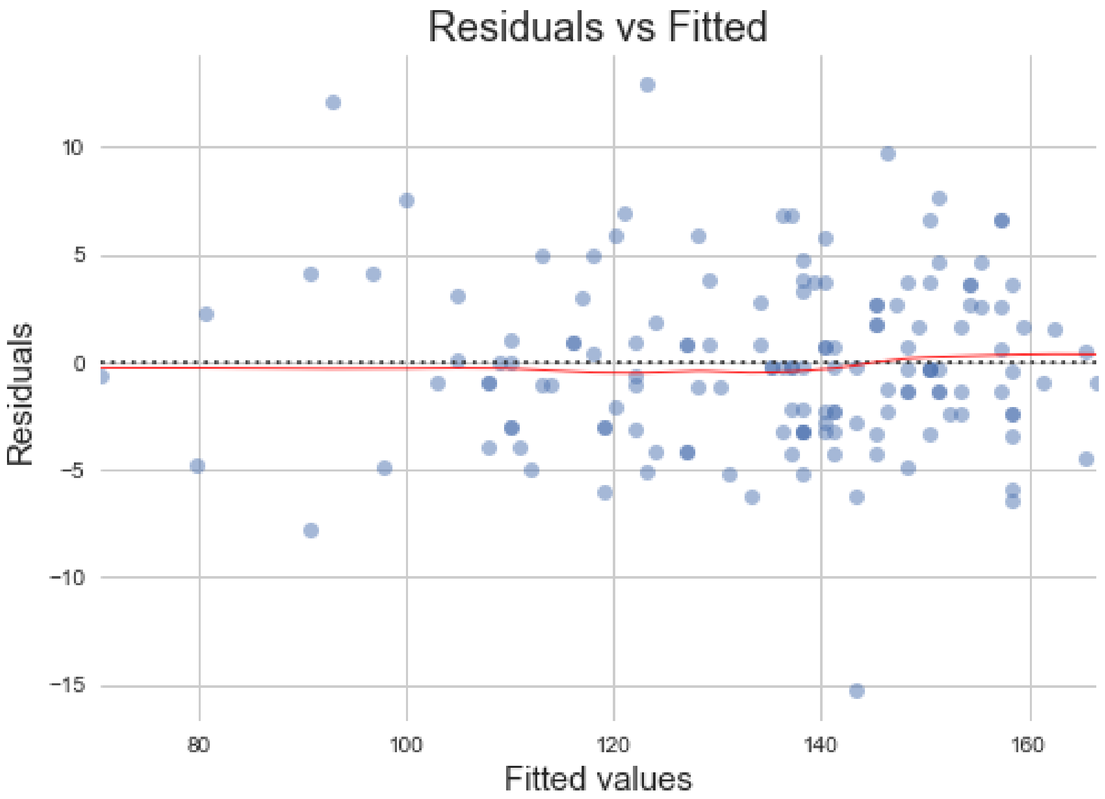


در این قسمت سه نمودار را با تابع ggplot رسم کردیم. در نمودار پراکنش یک رابطه خطی بین دو متغیر قد قبل و بعد از مصرف دارو مشاهده می‌شود.

```
ggplot(data) + aes(x="lastHeight") + geom_histogram(bins=10)
ggplot(data) + aes(x="Height",y="lastHeight") + geom_boxplot()
ggplot(data) + aes(x="Height",y="lastHeight") + geom_point()
```

نمودار های تشخیصی رگرسیون

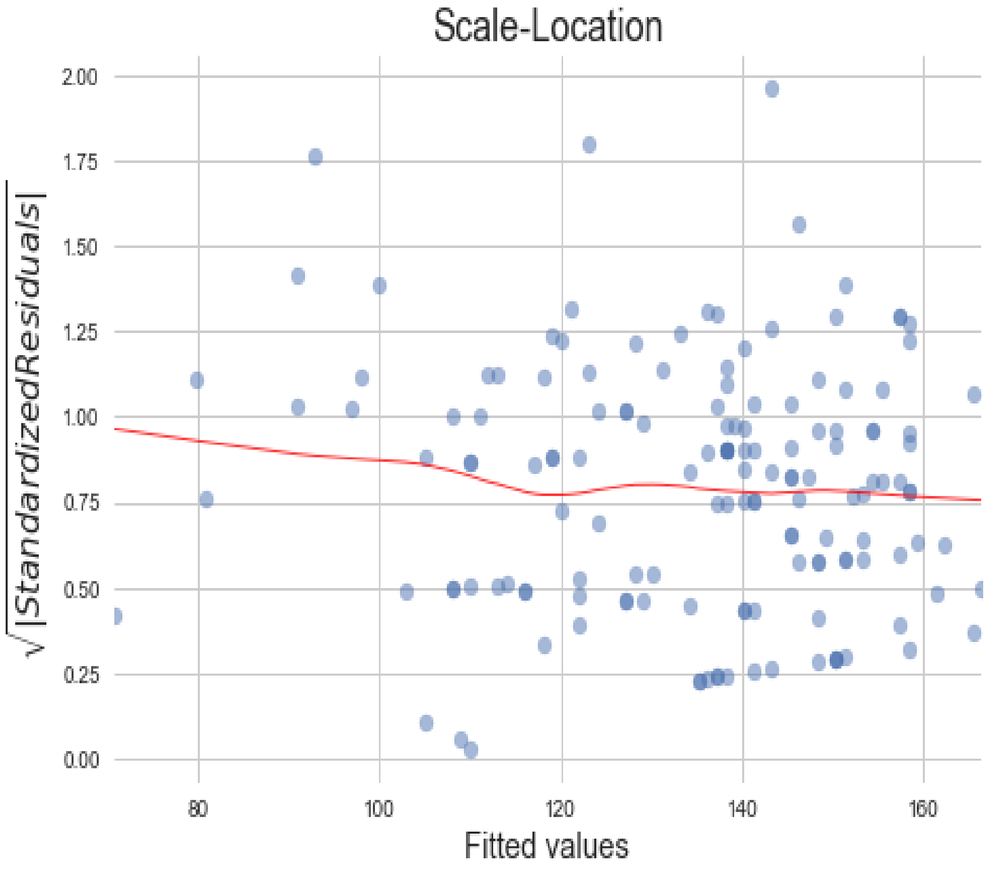
```
plot_lm_1 = plt.figure()
plot_lm_1.axes[0] = sns.residplot(model_fitted_y, data.columns[-1], data=data,
                          lowess=True,
                          scatter_kws={'alpha': 0.5},
                          line_kws={'color': 'red', 'lw': 1, 'alpha': 0.8})

plot_lm_1.axes[0].set_title('Residuals vs Fitted')
plot_lm_1.axes[0].set_xlabel('Fitted values')
plot_lm_1.axes[0].set_ylabel('Residuals')
```



در این قسمت نمودار مانده‌ها (جملات خطا) در مقابل مقادیر برازش داده شده متناظر (داده مربوط به قد کودکان) را رسم کردیم. این نمودار برای بررسی ثبات واریانس و مناسب بودن الگو استفاده می‌شود. اگر مدل برازش شده مناسب باشد این نمودار باید نسبت به نقطه ۰ متقارن باشد. با توجه به نمودار ثبات واریانس و مناسب بودن مدل تایید می‌شود.

نمودار های تشخیصی رگرسیون

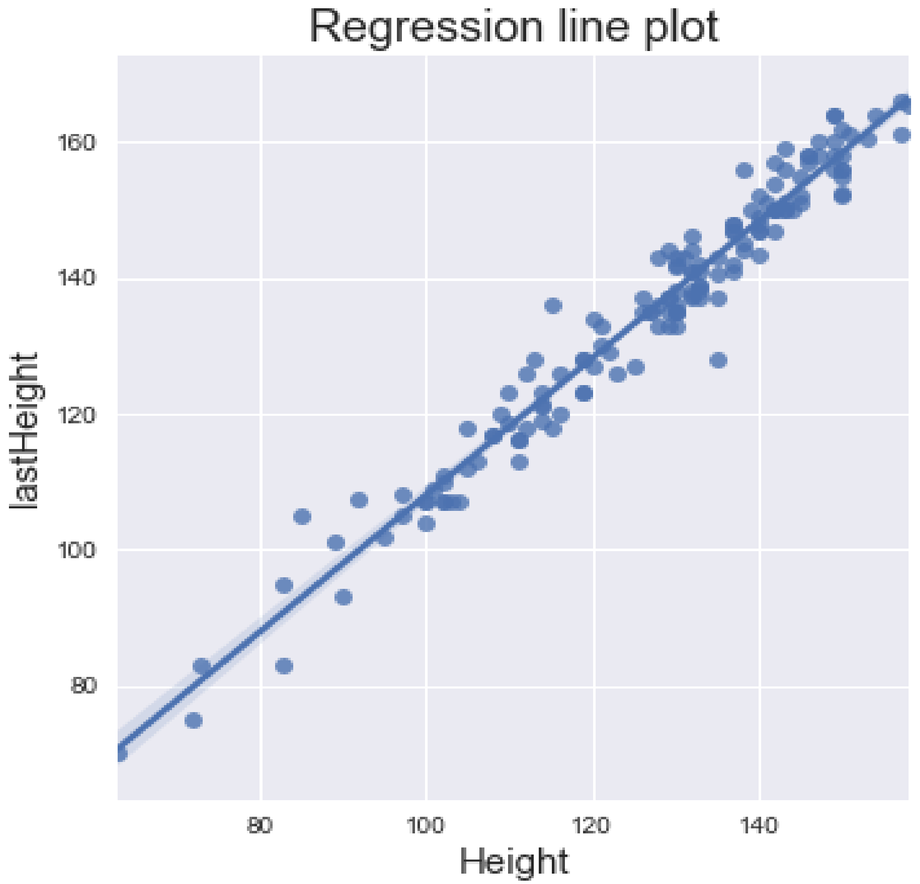
```
plot_lm_2 = plt.figure()
plt.scatter(model_fitted_y,model_norm_residuals_abs_sqrt,alpha=0.5)
sns.regplot(model_fitted_y,model_norm_residuals_abs_sqrt,
            scatter=False,
            ci=False,
            lowess=True,
            line_kws={'color': 'red', 'lw': 1, 'alpha': 0.8})
plot_lm_2.axes[0].set_title('Scale-Location')
plot_lm_2.axes[0].set_xlabel('Fitted values')
plot_lm_2.axes[0].set_ylabel('$\sqrt{|Standardized Residuals|}$')
```



در این قسمت نمودار مانده‌های استاندارد شده (جملات خطا استاندارد) در مقابل مقادیر برازش داده شده متناظر (داده مربوط به قد کودکان) را رسم کردیم. برای بررسی فرض ثابت بودن واریانس مانده‌ها، از این نمودار استفاده می‌شود. اگر مشاهدات در این نمودار، به صورتی باشند که شکل یک قیف افقی را نمایش دهند، مشخص است که واریانس مانده‌ها با افزایش مقادیر برازش شده، افزایش می‌یابد و فرض ثبات واریانس برای مانده‌ها را نمی‌توان تایید کرد. در این نمودار فرض ثبات واریانس تایید می‌شود.

نمودار های تشخیصی رگرسیون

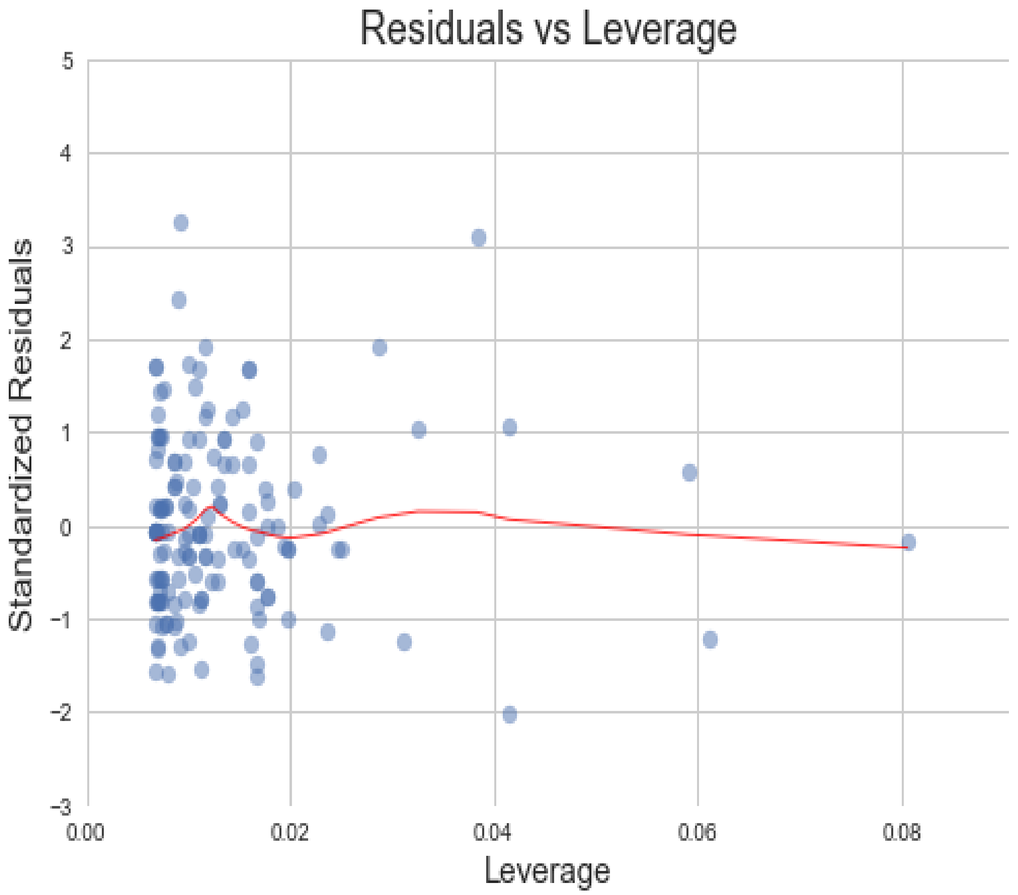
```
import seaborn as sns
ax=sns.lmplot(x ='Height', y ='lastHeight', data = data)
ax = plt.gca()
ax.set_title("Regression line plot")
```



در این قسمت نمودار خط رگرسیون برازش داده شده بر متغیر قد قبل از مصرف دارو در مقابل قد بعد از مصرف دارو را رسم کردیم. وجود یک رابطه خطی قوی مثبت بین دو متغیر مشاهده می‌شود و می‌توان گفت هرچه X افزایش می‌یابد Y نیز افزایش می‌یابد. (X و Y را در اسلاید فراخوانی داده‌ها مشخص کردیم)

نمودار های تشخیصی رگرسیون

```
plot_lm_3 = plt.figure()
plt.scatter(model_leverage, model_norm_residuals, alpha=0.5);
sns.regplot(model_leverage, model_norm_residuals,
            scatter=False,
            ci=False,
            lowess=True,
            line_kws={'color': 'red', 'lw': 1, 'alpha': 0.8});
plot_lm_3.axes[0].set_xlim(0, max(model_leverage)+0.01)
plot_lm_3.axes[0].set_ylim(-3, 5)
plot_lm_3.axes[0].set_title('Residuals vs Leverage')
plot_lm_3.axes[0].set_xlabel('Leverage')
plot_lm_3.axes[0].set_ylabel('Standardized Residuals')
```



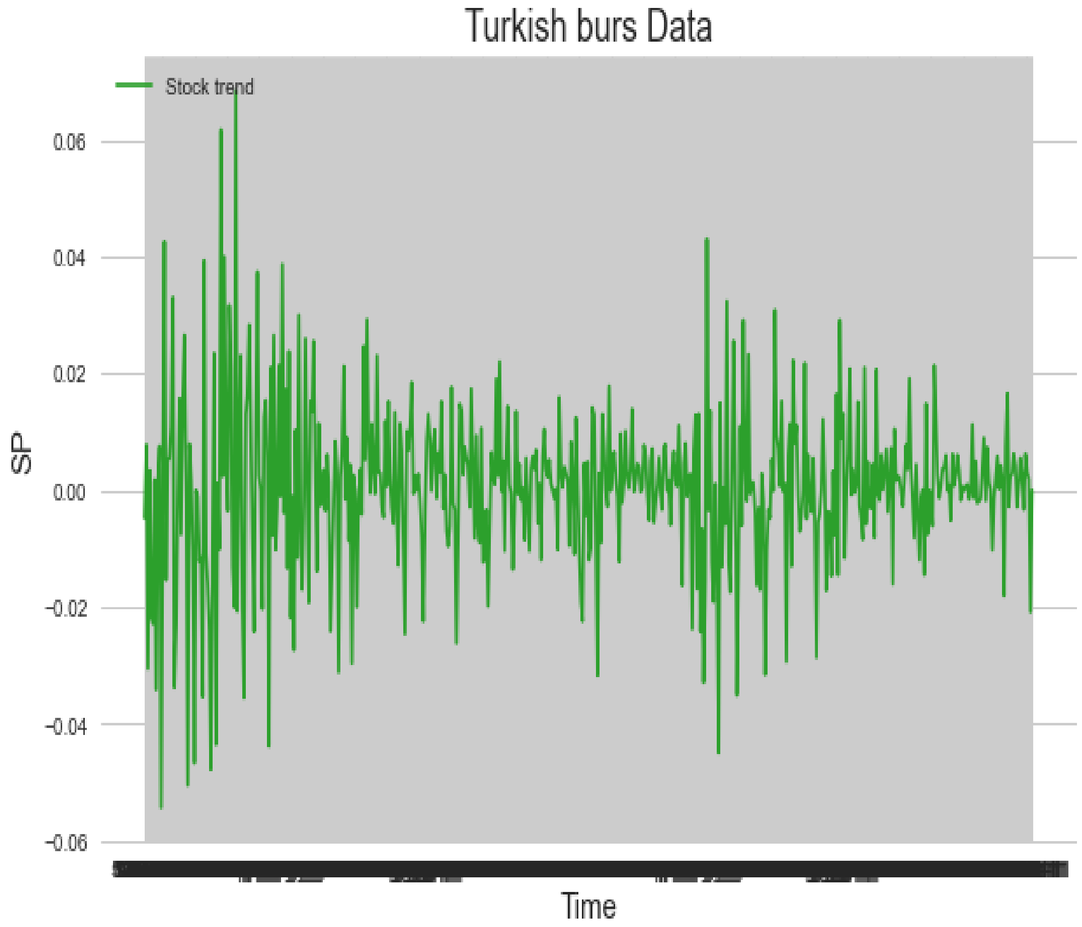
در این قسمت نمودار مانده‌های استاندارد شده (جملات خطا استاندارد) در مقابل مقادیر اهرمی (داده مربوط به قد کودکان) را رسم کردیم. در این نمودار به دنبال مشاهدات دورافتاده هستیم که باید مورد بررسی قرار داده شوند. داده دور افتاده داده‌هایی هستند که نسبت به سایر داده‌های در دست بررسی، تفاوت قابل ملاحظه‌ای داشته باشد و همچنین رفتار متفاوتی ارائه کنند. با توجه به نمودار داده‌ای که فاصله‌ی خیلی زیاد از دیگر مشاهدات داشته باشد، مشاهده نمی‌گردد.

نمودار سری‌زمانی

```
# time series plot #
import pandas as pd
data2=pd.read_csv(r"D:\statistical\trm 9\mohasebat amari\project\data_akbilgic.csv")
print(data2)

plt.rc('font', size=12)
fig, ax = plt.subplots(figsize=(10, 6))
ax.plot(data2.date, data2.SP, color='tab:green',label="Stock trend")
ax.set_xlabel('Time')
ax.set_ylabel('SP')
ax.set_title('Turkish burs Data')
ax.grid(True)
ax.legend(loc='upper left')
```

```
        date      SP
-5     0Jan-0.004679-09
-6     1Jan-09  0.007787
-7     2Jan-0.030469-09
-8     3Jan-09  0.003391
-9     4Jan-0.021533-09
          ...      ...       ..
-16  531Feb-11  0.006238
-17  532Feb-11  0.003071
-18  533Feb-11  0.001923
-21  534Feb-0.020742-11
-22  535Feb-11  0.000000

536]rows x 2 columns[
```

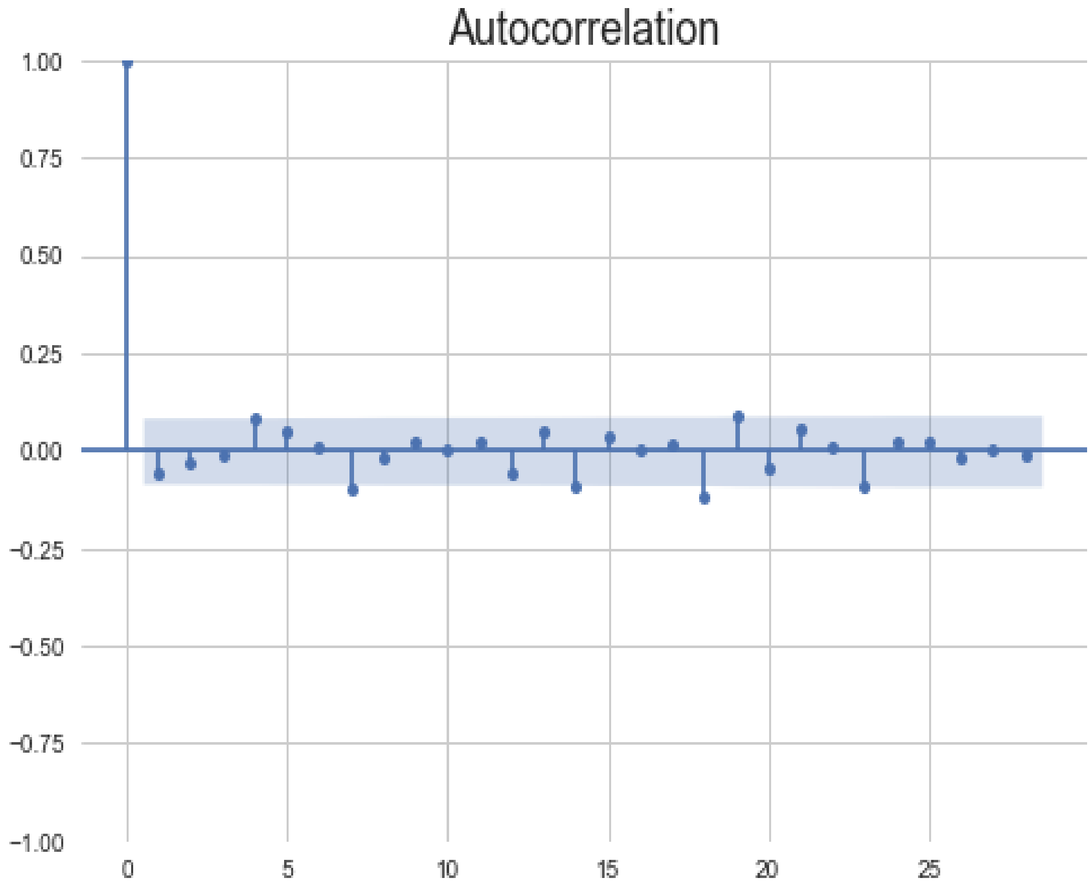
در این قسمت برای رسم نمودار های سری‌زمانی از داده های مربوط به سهام بورس کشور ترکیه استفاده کردیم که در طی سال های ۲۰۰۹ تا ۲۰۱۱ ثبت شده‌اند. در نمودار سمت راست سری‌زمانی متغیر SP در سه سال متوالی را مشاهده می‌کنید.



نمودار خودهمبستگی

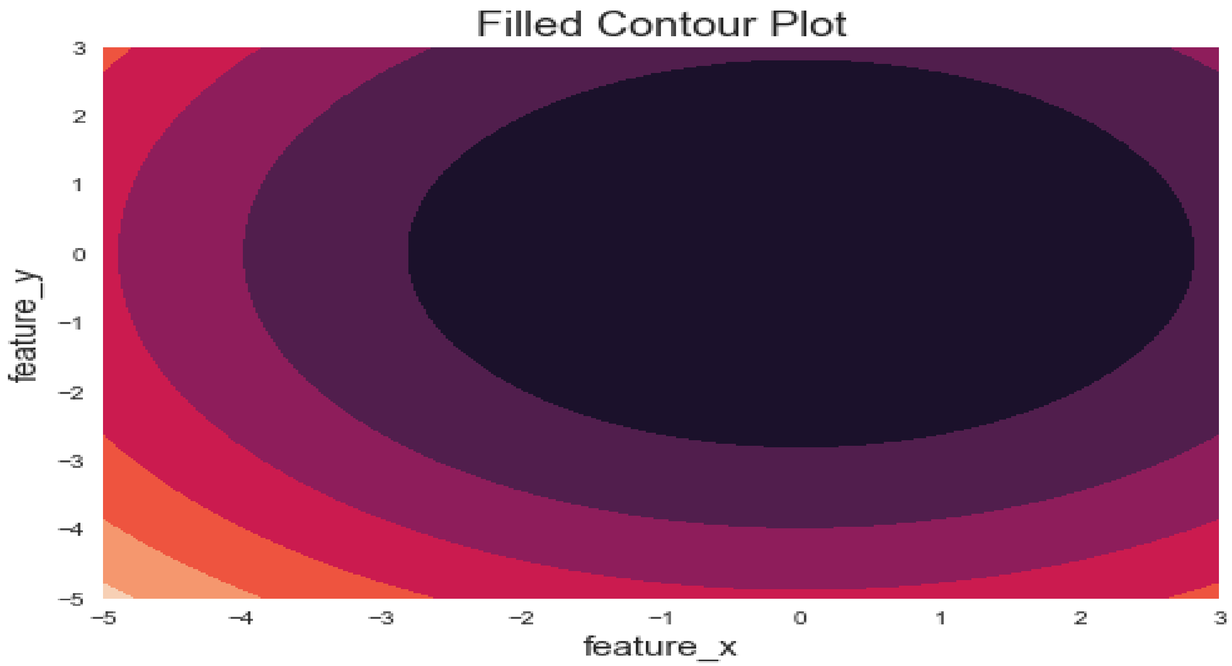
```
from statsmodels.graphics.tsaplots import plot_acf
plot_acf(data2.SP)
plt.show()
```

در این قسمت نمودار تابع خودهمبستگی را رسم کردیم. همانطور که مشاهده می‌شود ضریب خودهمبستگی برای تاخیر ۰ (که به معنی همبستگی سری با خودش است) برابر با ۱ محاسبه شده است. به ترتیب با افزایش تاخیرها نیز مقدار همبستگی بطور فاحش کاهش یافته است. بنابراین می‌توان نتیجه گرفت که داده‌ها خودهمبسته نیستند.



نمودار کانتور

```
# contour plot #
import matplotlib.pyplot as plt
import numpy as np
feature_x = np.linspace(-5.0, 3.0, 70)
feature_y = np.linspace(-5.0, 3.0, 70)
# Creating 2-D grid of features
[X, Y] = np.meshgrid(feature_x, feature_y)
fig, ax = plt.subplots(1, 1)
Z = X ** 2 + Y ** 2
# plots filled contour plot
ax.contourf(X, Y, Z)
ax.set_title('Filled Contour Plot')
ax.set_xlabel('feature_x')
ax.set_ylabel('feature_y')
plt.show()
```



در این قسمت نمودار کانتور را رسم کردیم. نمودار کانتور روشی است برای نمایش یک سطح سه بعدی در یک صفحه دو بعدی. در این نمودار هرچه به مناطق پر رنگ‌تر نزدیک شوید مقدار داده‌ای در آن منتطقه بیشتر است. به پررنگ ترین منطقه اصطلاحا نوک قله می‌گویند و هرچه از نوک قله دورتر شویم مقدار داده‌ای کاهش می‌یابد.

منابع

1. https://www.python.org/psf/
2. https://fileniko.com/entry/209712
3. شرکت داروسازی، داده‌های مربوط به رشد قد کودکان
4. https://www.machinelearningplus.com/plots/top-50-matplotlib-visualizations-the-master-plots-python/
5. https://archive.ics.uci.edu/ml/datasets.php، داده‌های مربوط به سری‌زمانی

با تشکر از توجه شما