Applied Data Science with Python

End-to-End Statistics Application with Python

Learning Objectives

By the end of this lesson, you will be able to:

- Perform basic statistical analysis on the sales data of a product and report the inference

- Reconstruct a bell curve using the SciPy library

- Predict the stock price based on the price variation for the past five years

Business Scenario

ABC Corporation's financial analyst wants to analyze revenue trends and forecast financial performance. To achieve this, conduct a statistical analysis of the sales data for a specific product or category. By examining historical sales data and calculating measures such as mean, median, standard deviation, and quartiles, one can gain insights into sales performance, including central tendency, variability, and distribution.

Using the SciPy library, reconstruct a bell curve to visualize the probability distribution of sales, enabling more accurate revenue forecasts. Armed with this information, the analyst can provide valuable insights to the management team, influencing strategic decision-making and resource allocation.

Business Scenario

Furthermore, the financial analyst is seeking to forecast the potential price movement of GOOG stock by analyzing its adjusted closing price over the past five years.

The challenge is to generate 50-day and 200-day moving averages (DMAs) and visualize the trends alongside actual prices. Also, exploratory data analysis is required to identify the best statistical features using OLS regression.

The final report will reveal correlations between the closing price and DMAs, as well as offer insights into the stock's future prospects. This analysis will be critical for investors and financial analysts who need to make informed investment decisions by comprehending GOOG's trends.

Basic Statistics with Python: Problem Statement

Basic Statistics with Python: Problem Statement

Consider pivot.csv, a product's sales data generated for September 2019.

There are two aspects of the product: units and sales.

Basic Statistics with Python: Problem Statement

Use the file as a basis to:

Perform basic statistical analysis and provide results

Use appropriate plots for displaying some of the statistical features

Group the analysis into two parts: weekly and monthly

Generate report on JupyterLab Notebook using markdown
text as well as code; plot and submit the report professionally

Analysis at the end is a must and the report should contain the inferences.

Assisted Practices

Let's understand the topic below using Jupyter Notebooks.

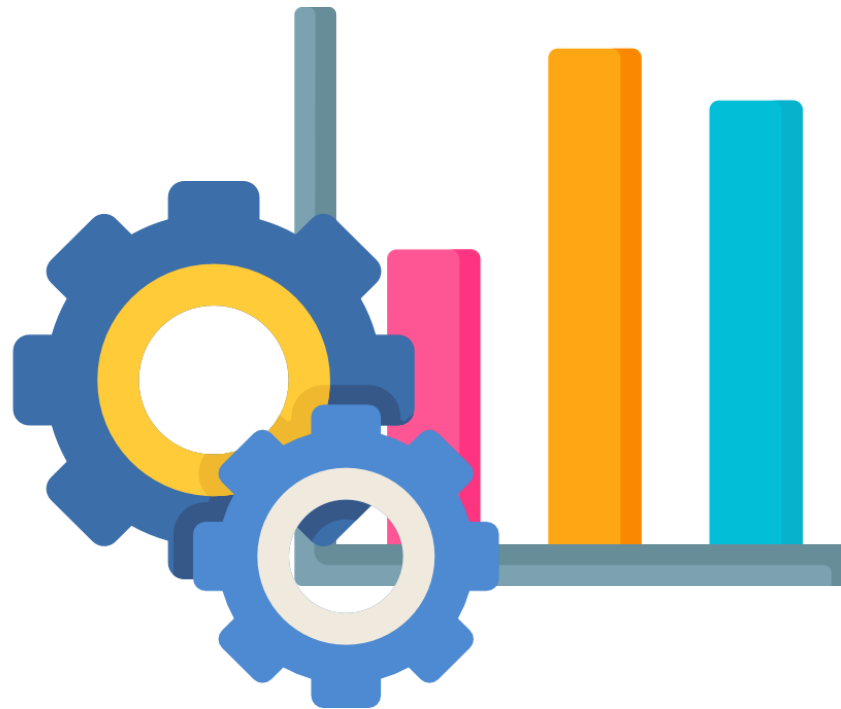- 13.3_Basic Statistics with Python: Solution

Note: Please download the pdf files for each topic mentioned above from the Reference Material section.

SciPy for Statistics: Problem Statement (Fixing a Bell Curve)

SciPy for Statistics: Problem Statement

Generate 1100 numbers, between 90 and 200, using NumPy's arange function.

This is a univariate problem.

SciPy for Statistics: Problem Statement

Assume that this represents the data on the height of a typical population in the 16–24 age group in Canada.

Perform mean, standard deviation, norm, and fit Bell Curve using SciPy libraries.

Use these libraries: Norm of scipy.stats and Statistics.

Generate a report on the JupyterLab Notebook using markdown text and code, and plot.

Analysis at the end is a must and the report should contain the inferences.

Let's understand the topic below using Jupyter Notebooks.

- 13.5_SciPy for Statistics: Solution

Note: Please download the pdf files for each topic mentioned above from the Reference Material section.

Advanced Statistics with Python: Problem Statement

Advanced Statistics with Python: Problem Statement

Find the possible futuristic price movement of the stock 'GOOG', based on its price over the past five years.

The analysis must be based on the adjusted closing price of the stock.

Advanced Statistics with Python: Problem Statement

Use the Yahoo yfinance package for the analysis to generate the following:

50-day moving average (50 DMA)

200-day moving average (200 DMA)

Plot the graph with the actual price.

Advanced Statistics with Python: Problem Statement

Use the following libraries:

yfinance

To get historical data

Ticker symbol 'GOOG'

To perform analysis

Generate the report in the JupyterLab Notebook using text and code, and plot.

Analysis at the end is a must and the report should contain inferences.

Let's understand the topic below using Jupyter Notebooks.

- 13.7_Advanced Statistics with Python: Solution

Note: Please download the pdf files for each topic mentioned above from the Reference Material section.

Key Takeaways

- There are two aspects of a product: units and sales.

- SciPy libraries can be used to perform the usual mean, standard deviation, norm, and fit the Bell Curve.

- JupyterLab Notebook can be used to generate report using the text and code.

Thank You