INFORMATICS PRACTICES PROJECT FILE

2024-25

CERTIFICATE

Done By: Prathamesh Mangnale

Co-Members: Sakshi Prabhakar Ingale

Manleen Kaur

CLASS: XII­­th B

This is to certify that this project entitled \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ has been submitted by\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ of

Class XII \_\_\_ bearing HTNO \_\_\_\_\_\_\_\_\_\_\_\_\_ in fulfillment of the course Informatics Practices of Class XII(CBSE Board) for the academic year 2020 - 2021 to the Department of Computer Science, Army Public School, Bolarum Secunderabad and it is carried out under my guidance and supervision.

**Internal External Principal**

ACKNOWLEDGEMENT

At the outset I am thankful to our school for giving me the opportunity to prepare the project report using Python Pandas.

I would like to express my gratitude to all my teachers and my laboratory instructor, who have constantly guided in the completion of the project work.

Thanking one and all

CONTENTS

S.No. Topic Page Numbers

01. System Requirements 4

02. Technology Used 5 - 9

a) Python

b) Pandas

c) DataFrames

d) CSV File

e) Visualizing Data & matplotlib

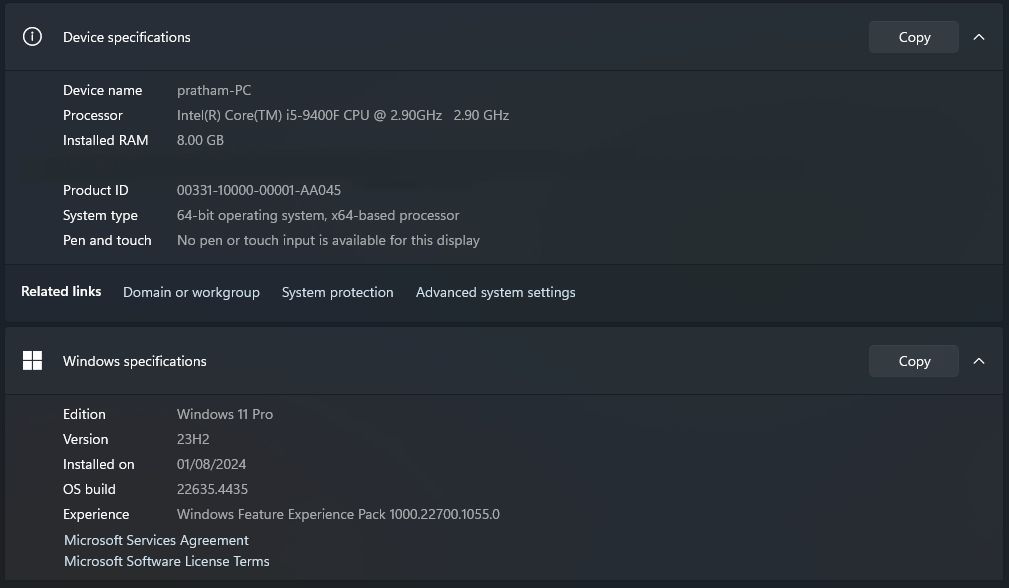
03. About The Project 10

04. Coding and Implementation 11 - 18

05. Bibliography 19

**SYSTEM REQUIREMENTS**

HARDWARE REQUIREMENTS:



SOFTWARE REQUIREMENTS:

1. Python IDLE 3.11
2. Microsoft Visual Studio Code
3. Microsoft Word

**TECHNOLOGY USED**

PYTHON:

Python Programming language was developed by Guido Van Rossum in February 1991.

Python is an easy to learn yet powerful object-oriented programming (OOP) language. It is not so high-level language as compared to C, C++ and Java etc.

pandas:

Pandas is the most popular library in the scientific Python ecosystem for doing data analysis, Pandas is capable of many tasks including:

* It can read or write in many different data formats (integer, float, double, etc.).
* It can calculate in all the possible ways data is organized i.e., across rows and down columns,
* It can easily select subsets of data from bulky data sets and even combine multiple datasets together.
* It has functionality to find and fill missing data.
* It allows you to apply operations to independent groups within the data.
* It supports reshaping of data into different forms.
* It supports advanced time-series functionality (Time series forecasting is the use of a model to predict future values based on previously observed values.)
* It supports visualization by integrating matplotlib and seaborn etc. libraries.

DataFrame:

A DataFrame is a Pandas structure, which stores data in two-dimensional way. It is actually a two-dimensional (tabular and spreadsheet like) labeled array, which is actually an ordered collection of columns where columns may store different types of data, e.g., numeric or string or floating point or Boolean type etc.

Major characteristics of a DataFrame are :

* It has two indexes or we can say that two axes - a row index (axis = 0) and a column index (axis = 1).
* Conceptually it is like a spreadsheet where each value is

identifiable with the combination of row index and column index. The row index is known as index in general and the column index is called the column-name.

* The indexes can be of numbers or letters or strings.
* There is no condition of having all data of same type across columns; its columns can have data of different types.
* DataFrames are value-mutable (its values can be easily changed)
* DataFrames are size-mutable (rows/columns can be easily added or deleted)

CSV FILE:

Refers to the tabular data saved as plaintext where data values are separated by commas.

The CSV format is popular as it offers following advantages :

* A simple, compact and ubiquitous format for data storage.
* A common format for data interchange.
* It can be opened in popular spreadsheet packages like MS-Excel, Calc etc.
* Nearly all spreadsheets and databases support import /export to csv format.

Python's Pandas library offers two functions read\_csv() and to\_csv( ) that help you bring data from a CSV file into a DataFrame and write a data frame’s data to a CSV file.

VISUALIZING DATA & matplotlib:

Data Visualization basically refers to the graphical and visual representation of information and data using visual elements like charts, graphs, and maps, etc.

Data Visualization is immensely useful in decision making unveils patterns, trends, outliners, correlations etc. in the data helping decision-makers understand the meaning of data to drive decisions.

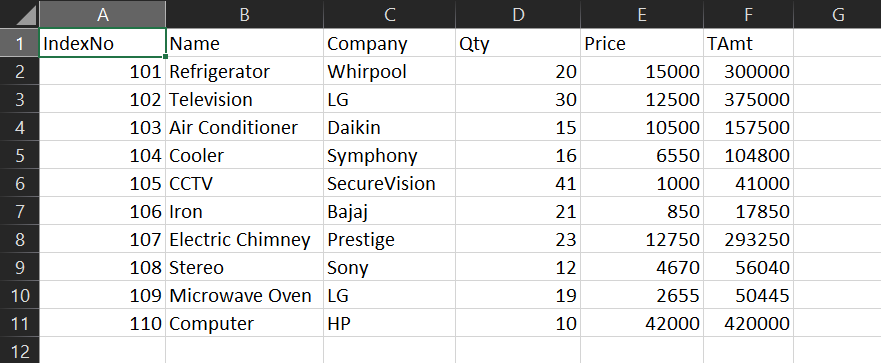
The matplotlib is a Python library that provides many interfaces and functionalities for 2D graphics. matplotlib is a high-quality plotting library of python that provides both very quick way to visualize data from Python and publication-quality figures in many formats.

The matplotlib library offers many different named collections of methods; PyPlot is one of such interfaces, a collection of methods within matplotlib which allows used to construct 2D plots easily and interactively.

**ABOUT THE PROJECT**

**CODING AND IMPLEMENTATION**

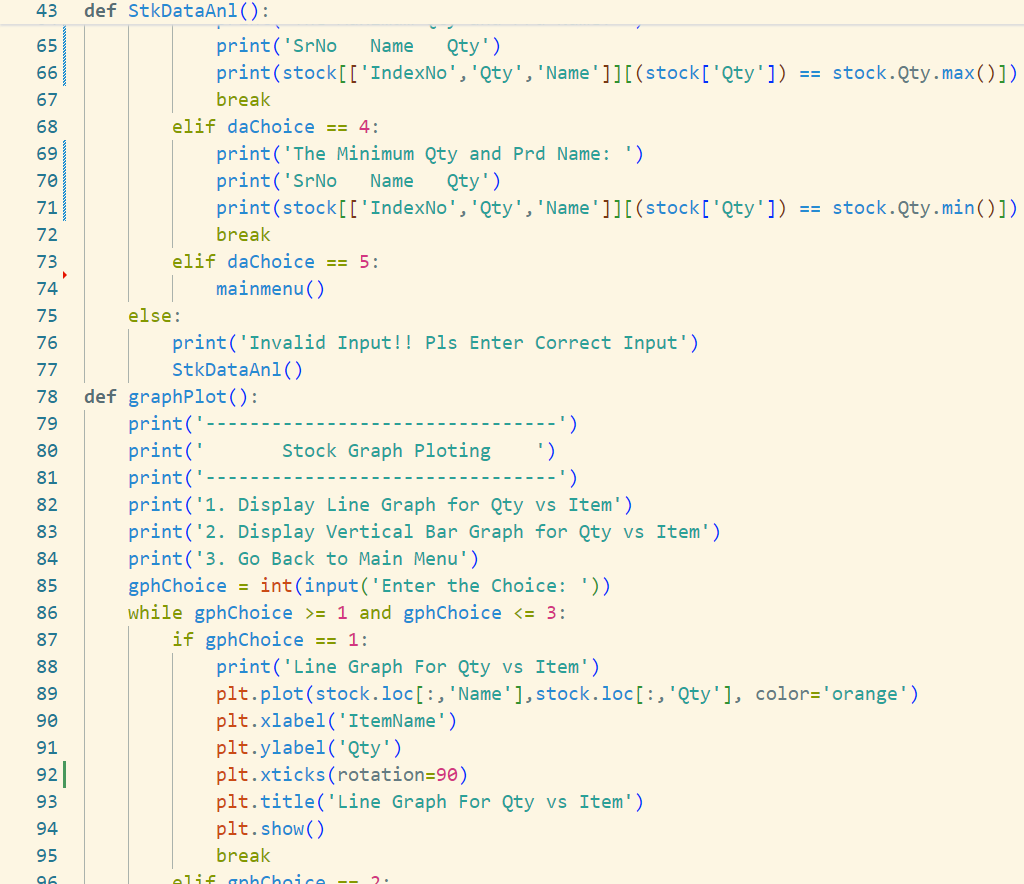
* EXCEL SHEET (.csv File):

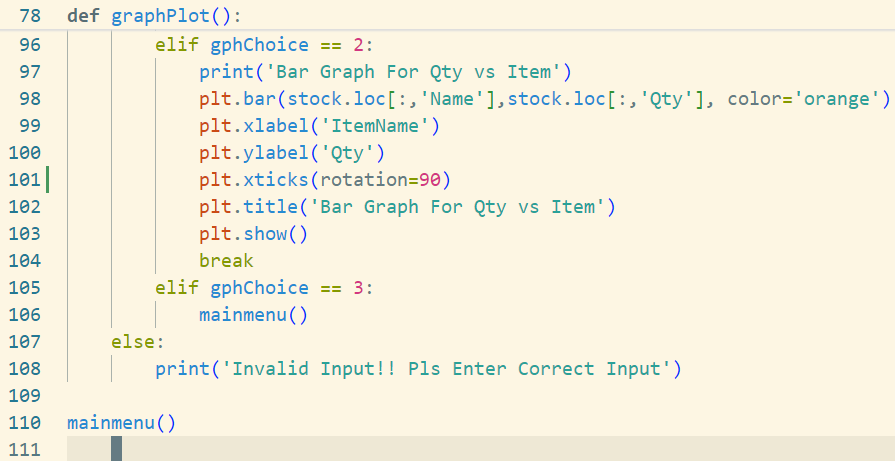


* CODING IN PYTHON (.py File):

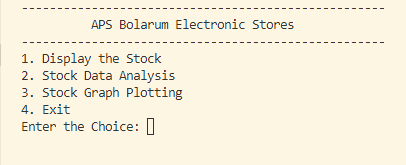


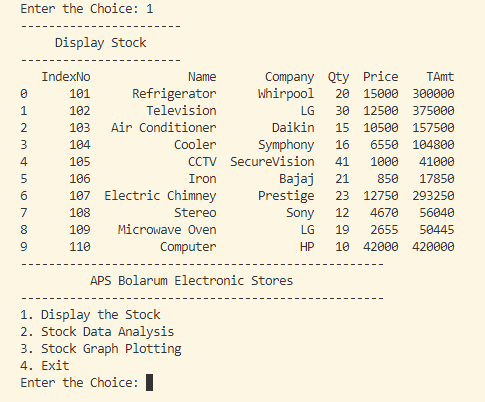


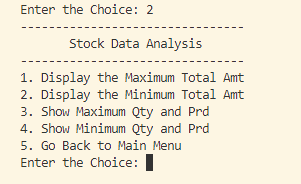


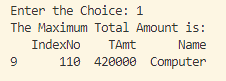


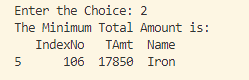
* OUTPUT WINDOW:

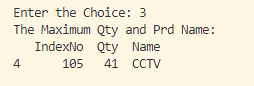


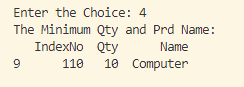


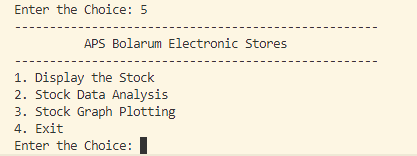


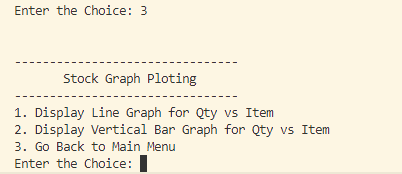


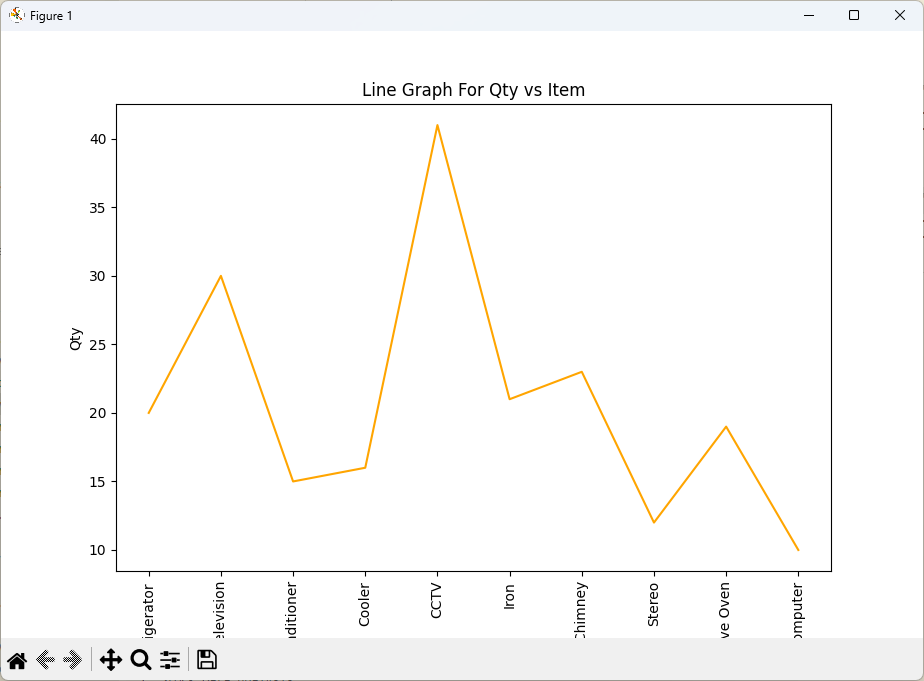


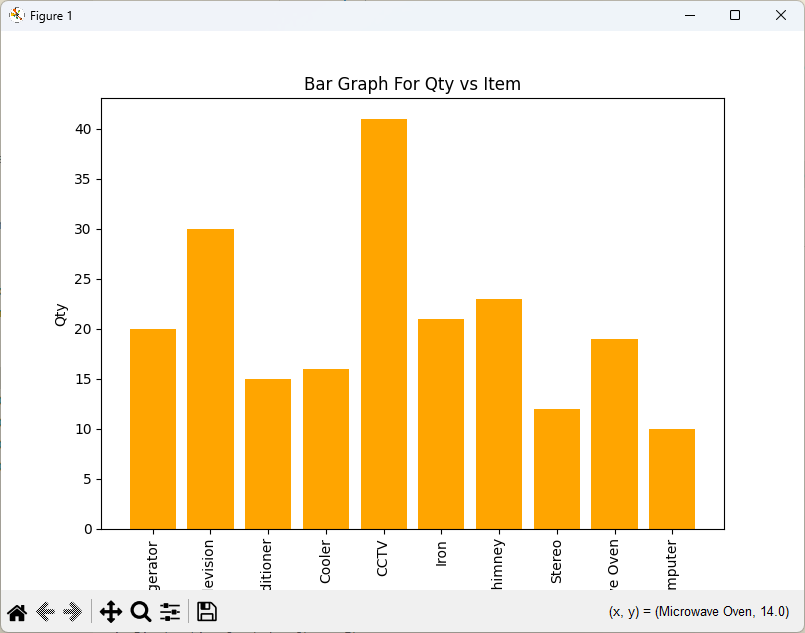


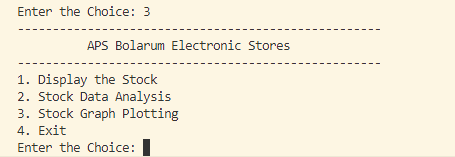














**BIBLIOGRAPHY**

* Class 12th NCERT Textbook
* E – Resources
* [www.stacloverflow.com](http://www.stacloverflow.com)