

---

# PYTHON DATA SCIENCE

---

Project 1

DocID	Version	Language	Date	Page
PY/DS/PR01	1.0	En	19 <sup>th</sup> Sept 2024	1

## Contents

Instructions before the project execution .....	2
Project Overview .....	3
Python Libraries for Data Science .....	3
NumPy .....	3
Pandas.....	3
Matplotlib and Seaborn .....	3
Data Manipulation and Analysis with Pandas .....	3
Importing and Exporting Data .....	3
Data Cleaning and Preparation.....	3
Data Transformation .....	3
Data Aggregation and Grouping .....	3
Exploratory Data Analysis (EDA) .....	4
Data Visualization Techniques .....	4
Importance of Data Visualization .....	4
Types of Visualizations .....	4
Basic Plotting with Matplotlib .....	4
Advanced Visualization with Seaborn .....	4
Interactive Visualization with Plotly .....	4
Basic Chart Types .....	4
Bar Charts.....	4
Line Charts.....	4
Pie and Donut Charts.....	4
Analysis to be done. ....	5
Descriptive Statistics: .....	5
Trend Analysis: .....	5
Customer Segmentation:.....	5
Account Type Analysis: .....	5
Transaction Type Analysis: .....	5
City-wise Analysis: .....	5
Time-based Insights: .....	5
Correlation Analysis: .....	5

DocID	Version	Language	Date	Page
PY/DS/PR01	1.0	En	19 <sup>th</sup> Sept 2024	2

## Instructions before the project execution

- READ THE ENTIRE DOCUMENT BEFORE BEGINNING
- Write down the understanding
- PLAN properly start to end the path of working on this project
- Do not use any code generation tools for completion
- In the project review, questions about any of the above can be asked, including code, logic and the reasoning. hence it is advised to do the project independently.
- Document the project execution from start to finish

DocID	Version	Language	Date	Page
PY/DS/PR01	1.0	En	19 <sup>th</sup> Sept 2024	3

## Project Overview

The goal of this project is to give you hands-on experience in applying Python-based data science techniques using libraries like Pandas, NumPy, Matplotlib, and Seaborn, with a dataset of 10,000 financial transactions. You will explore various aspects of data manipulation, cleaning, and visualization, with Python, pandas, matplotlib, numpy and seaborn to achieve the results.

---

## Python Libraries for Data Science

### NumPy

- Array creation, indexing, and manipulation
- Basic statistical operations (mean, sum, std)

### Pandas

- DataFrames and Series objects
- Loading the financial transactions dataset
- Viewing and summarizing data

### Matplotlib and Seaborn

- Basic plotting (bar charts, line charts, etc.)
  - Customizing visualizations (labels, colors, grids)
- 

## Data Manipulation and Analysis with Pandas

### Importing and Exporting Data

- Loading .csv financial data

### Data Cleaning and Preparation

- Handling missing values
- Dropping unnecessary columns

### Data Transformation

- Changing data types, adding new columns
- Working with date/time formats in financial transactions

### Data Aggregation and Grouping

- Grouping by categories (e.g., transaction types)

Copyright © Kiran VVN. 2000-2024, All rights reserved

No part of this document shall be reused, copied in any form either in full/partial without exclusive written permission from kvvn@me.com

DocID	Version	Language	Date	Page
PY/DS/PR01	1.0	En	19 <sup>th</sup> Sept 2024	4

- Aggregating data to compute statistics like sum, count

## Exploratory Data Analysis (EDA)

- Descriptive statistics (mean, median, mode)
  - Visualizing distributions, trends, and correlations
- 

# Data Visualization Techniques

## Importance of Data Visualization

- Why visual representation matters in data analysis

## Types of Visualizations

- Line charts, bar charts, histograms, scatter plots

## Basic Plotting with Matplotlib

- Line charts for time-based transaction data
- Bar charts to compare transaction types

## Advanced Visualization with Seaborn

- Heatmaps, pair plots, correlation matrices
- Customizing visual aesthetics (themes, palettes)

## Interactive Visualization with Plotly

- Line and scatter plots with tooltips and zoom
- 

# Basic Chart Types

## Bar Charts

- Visualizing transaction types by volume
- Grouped and stacked bars for comparing categories

## Line Charts

- Plotting transaction amounts over time

## Pie and Donut Charts

- Visualizing categorical distributions (e.g., payment methods)

Copyright © Kiran VVN. 2000-2024, All rights reserved

No part of this document shall be reused, copied in any form either in full/partial without exclusive written permission from kvvn@me.com

DocID	Version	Language	Date	Page
PY/DS/PR01	1.0	En	19 <sup>th</sup> Sept 2024	5

## Analysis to be done.

### Descriptive Statistics:

- Average transaction amount by city, account type, or transaction type.
- Total number of transactions per customer or per city.
- Distribution of transaction amounts (mean, median, mode).

### Trend Analysis:

- Time-series analysis of transactions over the last two years.
- Monthly or quarterly trends of deposits and withdrawals.

### Customer Segmentation:

- Group customers based on transaction amounts (e.g., high-value vs. low-value customers).
- Analyze behavior by customer location (e.g., average transaction amount per city).

### Account Type Analysis:

- Compare transaction amounts and frequency across checking, savings, and credit accounts.

### Transaction Type Analysis:

- Breakdown of transactions by type (e.g., deposit vs. withdrawal vs. transfer).
- Analyze patterns in transaction types over time.

### City-wise Analysis:

- Identify cities with the highest volume of transactions.
- Compare average transaction sizes across different cities.

### Time-based Insights:

- Analyze peak transaction times (e.g., by hour of the day, day of the week, or month).
- Seasonal patterns in spending or saving behavior.

### Correlation Analysis:

- Correlate transaction amounts with customer location or account type.
- Analyze relationships between different variables like account type and transaction type.