**Shiva Kumar Peddapuram ||** [**skp5.aiml@gmail.com**](mailto:skp5.aiml@gmail.com) **|| +1 2342962470**

Dataset: I have used dataset from Kaggle it is a open source software

<https://www.kaggle.com/datasets/shriyashjagtap/indian-personal-finance-and-spending-habits>

Below operations are done in Excel.

Before performing any operations on the dataset, it would be better to make a copy of the same dataset and store it in the new sheet so that makes easy for us to retrieve if in case any mistakes happen.

After making a copy of the dataset I will make sure to fix the top columns as freeze so that way it gets easy for me to analyse when I go to the last row.

So as a part of the next step I am converting all the columns into their respective formats by following the below instructions

Select the whole row and press right click on the mouse and then click on format cells then select the currency (As this dataset belongs to the Indian family income, I am converting the whole numbers into the India Currency).

As a next step I am also converting the currency format for the whole dataset it can be either done in two ways by selecting individual rows and then following the above instructions or we can select all the data which are needed to be converted into the currency and then press right click on the mouse courser and then select format and then select currency and make sure to keep the decimal point as 2 and then press ok.

Now save the whole dataset by pressing ctrl +s or by directly selecting save option on the excel.

Till here I have used excel.

As a next step I am using Jupiter notebooks as interface and then I am using Python as my Programming language to make sure whether the given set is perfect to do all the operations like checking all the null values, distinct rows and columns, etc.

Initially I have loaded all the necessary projects that are needed for performing the operations. As a next step I have started reading the csv file (The dataset which we have performed few operations earlier in MS-Excel.)

Next I have just gave a small try to see the edited dataset by using the python function head() and tail () and then I have seen the whole summary of the given datasets just to make sure all the columns are been edited.

Next step is too crucial fore the whole process that is checking whether the datasets has the null values or not if we have any niull values then we have several methods to rectify these situation few situations are by replacing the null value with zero the another option is by the mean of the whole column at the null values. So sometimes some experts uses standard deviatrion in the place of null value.

To our surprise we don’t have any nul;;l values in the given datasets that helps a lot in next steps.

So, for the next steps I am loading the following dataset in powerBI or Tableau for performing the visualizations

I am making sure to use all the tools such a way I can showcase all my skills I have gained during my masters.

Based on the provided deliverables and your content, here’s a structured and comprehensive document that aligns with the Data Analyst role. I’ll include each of the relevant sections as they apply to the work you've described.

---

# Data Analysis Documentation for [Project/Assignment Title]

## 1. Business Requirements Document (BRD)

### Objective

To prepare a clean, consistent dataset representing Indian family income data to enable meaningful analysis and derive actionable insights.

### Requirements

- Ensure data integrity and consistency.

- Format financial data according to Indian currency standards.

- Validate dataset quality using Python and Jupyter Notebooks to check for null values, unique entries, and data conformity.

---

## 2. System Requirements Specification (SRS)

### Software Requirements

- \*\*Excel:\*\* For initial data preparation and formatting.

- \*\*Python (Jupyter Notebook):\*\* For data validation, analysis, and quality assurance.

### Functional Requirements

- \*\*Data Backup:\*\* Create a duplicate of the original dataset to maintain a version history.

- \*\*Data Formatting:\*\* Apply currency and numerical formatting to match Indian Rupee standards.

- \*\*Data Validation:\*\* Use Python libraries to validate the dataset by identifying null values, distinct rows, and columns.

---

## 3. Business Analysis Plan

### Overview

- Use Excel for initial data cleaning and formatting to streamline the dataset.

- Employ Python for further analysis to ensure data accuracy before conducting statistical analysis.

- Prepare documentation to capture data processing steps, allowing for reproducibility.

### Analysis Tasks

1. Freeze header row in Excel for easy navigation.

2. Convert relevant data to currency format.

3. Validate dataset integrity using Python.

---

## 4. Business Case Document

### Justification

A clean and well-structured dataset will allow for reliable statistical analysis and pattern recognition, enabling the identification of key trends in family income data in India.

---

## 5. Gap Analysis Document

### Identified Gaps

- \*\*Data Consistency:\*\* Ensuring all currency values are standardized.

- \*\*Missing Values:\*\* Detecting and addressing null values in the dataset.

- \*\*Formatting Gaps:\*\* Some values may need adjustment to align with data analysis standards.

---

## 6. Stakeholder Management Plan

### Stakeholders

- \*\*Data Analysis Team:\*\* Responsible for data preparation, formatting, and validation.

- \*\*Business Unit:\*\* Requires accurate data for income pattern analysis.

- \*\*Technical Team:\*\* Supports in using Jupyter Notebook and Excel for data processing.

---

## 7. Requirements Traceability Matrix

| Requirement | Source | Status | Remarks |

|-------------------------|---------|------------|----------------------------------|

| Data Backup | Analyst | Completed | Duplicate created in Excel |

| Currency Formatting | Analyst | Completed | Set to Indian Rupee with 2 decimals |

| Data Validation in Python | Analyst | In Progress | Conducted in Jupyter Notebook |

---

## 8. Process Flow Diagram

Include a diagram to represent the steps taken from data preparation in Excel to validation in Python.

---

## 9. User Stories and Use Cases

### User Story

As a Data Analyst, I need to prepare and validate income data to ensure it is clean and formatted correctly so I can derive accurate insights.

### Use Case

1. User opens the dataset in Excel and creates a backup.

2. User formats currency and ensures headers are frozen.

3. User saves and moves to Jupyter Notebook for validation.

---

## 10. Functional Decomposition Document

1. \*\*Data Duplication:\*\* Create a backup of the dataset.

2. \*\*Currency Formatting:\*\* Standardize all financial data.

3. \*\*Validation:\*\* Use Python to check for nulls, unique values, and consistency.

---

## 11. Impact Analysis Report

If the dataset is not formatted and validated correctly, it may lead to inaccuracies in trend analysis and predictive modeling.

---

## 12. System Test Plan

1. \*\*Objective:\*\* Ensure all formatting and validation processes meet data standards.

2. \*\*Test Environment:\*\* Excel for formatting, Jupyter Notebook for validation.

3. \*\*Test Cases:\*\* Verify null values, ensure consistent currency format, and confirm uniqueness of entries.

---

## 13. Test Cases

| Test Case | Description | Expected Result |

|-----------------------------|----------------------------------------------------|----------------------------------|

| Freeze Header Row | Check if header row remains visible when scrolling | Header remains fixed |

| Currency Format Verification | Ensure all currency values display as INR | Consistent currency formatting |

| Null Value Check | Verify no null values remain | Dataset free of null values |

---

## 14. UAT Progress Report

\*\*Status:\*\* Formatting and initial validation completed. Currently progressing through Python-based validation steps.

---

## 15. Statistical Analysis Reports

Upon completion of validation, generate a statistical analysis report to summarize key income trends, average incomes, and any detected patterns.

---

## 16. Cleaned and Validated Datasets

The final dataset will be saved in a separate file with clear documentation on cleaning steps.

---

## 17. Data Quality Assessment

- \*\*Null Values:\*\* Checked and addressed.

- \*\*Currency Format:\*\* Standardized across all income columns.

- \*\*Duplicate Rows:\*\* Removed if detected.

---

## 18. Data Visualization Reports

Following validation, create data visualization reports to highlight key insights and income trends using Excel charts or a data visualization tool (e.g., Power BI, Tableau).

---

## 19. Business Intelligence Presentation

Summarize the analysis and findings in a presentation to communicate key patterns and trends to stakeholders.

---

This document should serve as a comprehensive guide for the data preparation, analysis, and documentation stages. Each section aligns with the requirements for a Data Analyst role and ensures that the process is well-documented, structured, and ready for further analytical tasks. Let me know if you’d like further customization or specific details in any section.