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

**Functional Deliver abilities:**

1. **Track Daily Expenses:** The dataset allows tracking of daily expenses, giving extra information to monitor users spending habits effectively.
2. **Categorize Transactions:** The one more key feature in the given dataset is they are organized into categories making easier to perform lot of operations and analyse spending patterns.
3. **Generate Basic Spending Reports**: Basic reports summarizing spending by category or time period can be generated from the dataset.
4. **Visualize Monthly Spending Patterns:** The dataset’s structure facilitates monthly spending pattern visualizations, providing insights into budgeting and spending behaviour.

**Documentation Deliverables:**

1. **Business Requirements Document (BRD):**
2. **Project Goals:** Personal Expenses Tracker.
3. **Objectives:** Creating multiple visualizations by taking KPI (Key Performing Indicators) into considerations like: (Personal expenses tracking, categorizing transactions based upon spendings, monitoring spending habits, analysing financial patterns.)
4. **High – Level requirements:** high knowledge on selection of datasets, Hands on experience in cleaning datasets, Expertise in Exploratory Data Analysis Process or Export Transform Load Process, understanding of several python libraries, Tableau, PowerBI, MS Word (Documentation).
5. **Dataset requirements:** There are several websites for the dataset collection I have used the dataset from Kaggle because it has several columns which might help me for making KPI Points more.
6. **Analytical Requirements:** For the analytical tools, I have used Python and Tableau. In few Cases we can also use PowerBI too.
7. **System Requirements Specification (SRS):**
8. **System Requirements:**

**Hardware Requirements:**

**Processor:** Intel core i5 (or Greater),

**Ram:** Minimum 8GB or higher (For handling larger datasets and for multitasking),

**Storage**: At least 15 GB Free disk space for dataset storage, temporary files, environments, documentation, visualization.

1. **Software Requirements**

**Operating System:** Windows 10 or new versions

1. **Dataset:** Kaggle account
2. **Coding Requirements:**

**Coding Language:** Python, R (Can use accordingly)

**Environment:** Jupyter Notebook or Jupyter labs installed (through Anaconda)

**Python Libraries:** Pandas, NumPy, matplotlib seaborn plotly

1. **Microsoft Excel:** Version 2016 for performing ET: (Extract, Transform, Load) processes that required data cleaning and reformation
2. **Documentation Tools:** Microsoft Word 2016 or greater, Google Docs.
3. **Additional Software’s:** GitHub (Postings the work updates into it),
4. **Visualization Tools:** PowerBI, Tableau
5. **Business Analysis Plan**

**Data Selection:**

Dataset: Selecting a dataset that requires all of our KPI (Key Performing indicators) requirement is bit complicated. Even I faced lot of challenges in selecting the dataset. Earlier I have selected lot of datasets, but they were not meeting my requirements. I have referred to several sources like (Government sites, Kaggle). Finally, I have used dataset from Kaggle as it is an open source where all the datasets are available for free. Below I am pasting the link to the dataset: Indian Personal Finance and Spending Habits Dataset – Kaggle

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

**Transformation:**

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.

**Tools Used:** MS-Excel, Kaggle

**Loading the dataset:**

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 being edited.

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

To our surprise we don’t have any null 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

**Tools Used:** Excel, Python in Jupyter Notebook,

**Visualizations:**

According to the KPI Performed several visualizations and noted few of the outcomes

From the visualization I have made the following key points that might help me to identify further results:

1. Most of the people with age group 18 has started working and gaining money in form of salaries,
2. Tier 2 people has most of the money generated
3. Out of those tiers 2 most of the people are self-employed and students.
4. Out of these tier 2 people spent most of their income on their grocery around (10,000 inr)
5. **Business Case Document:**

Selecting an appropriate dataset is a critical step in data analysis, as it directly impacts the insights and visualizations that can be generated. During my search, I have evaluated more than 3-5 datasets but encountered several challenges that limited their effectiveness for KPI-based analysis. Key issues included:

1. **Limited Attributes:** Some datasets contained too few attributes, restricting the depth and variety of KPIs and visualizations possible. This lack of dimensionality limited exploratory analysis and hindered meaningful insights.
2. **High Null Value Columns:** Several datasets had entire columns filled with null values, making it challenging to conduct thorough analysis or leverage certain attributes for KPIs without extensive data cleaning or imputation.
3. **Inconsistent Data Types:** In some cases, datasets lacked standardized data types, which complicated processing and led to additional steps for conversion or correction.
4. **Data Type Errors in Categorical Fields:** Certain categorical columns contained string values where numerical data was expected, resulting in errors during analysis and requiring data type corrections.
5. **Date Format Issues:** Irregular or incorrect date formats in some datasets posed challenges for time-series analysis and visualization, necessitating additional preprocessing steps to standardize date formats.

After evaluating these datasets, I selected the current dataset as it addressed these issues effectively. It contains a large number of attributes, has no null values, consistent data types, and is well-suited for comprehensive KPI analysis and visualization.

1. **Gap Analysis Document:**

Selecting a suitable dataset that suits all of our requirements which mainly includes KPI, Null Values, and few more factors that helps in deciding whether the selected datasets is suitable for the analysis or not.

I have seen several datasets for analysing the KPI but unable to find them after a long gap found the above dataset that helped me a lot.

**Analysis Plan:**

Gathering the requirements

Loading the dataset and Performing data Cleaning

Data Set Collection

Performing Python Operations

Making Dashboard in visualization tools

Performing Visualizations

Drawing Conclusion

**Visualization:**

Count of age v/s age

In the below visualization I have made the plot in between age and count of people.

And the outcome for the below visualization was in between the age of 18 to 50 Years.

A graph of a bar graph

Description automatically generated

In the below Image it depicts the tier wise occupation along with their income count.

In the below visualization we can clearly see that tier-2 people has more income.

A screenshot of a graph

Description automatically generated

Tierwise Income:

A graph of a number of people

Description automatically generated with medium confidence

Dependents breakdown Tier Wise:

In Tier-2 Dependents are more.

A screenshot of a computer

Description automatically generated

Final Dashboard

A screenshot of a computer

Description automatically generated