

Indian Personal Finance and Spending Habits

1.Dataset Description

Source:

- The dataset is available on **Kaggle** as *Indian Personal Finance and Spending Habits*.
- It contains data for about **20,000 individuals** capturing demographic, income, expense,
- The file is typically provided as data.csv (≈ 8.63 MB) in the Kaggle dataset.

Columns :

Column	Description
age	Age of the individual
gender	Gender of the individual (Male / Female / Other)
marital_status	Married / Unmarried / Other
occupation	Type of occupation (e.g. Salaried, Self-employed, Business, etc.)
income	Total monthly or annual income of the individual
dependents	Number of dependents in the household
monthly_expenses	Sum of all expense categories (or possibly broken into subcategories)
rent	Monthly rent (if applicable)
utilities	Utilities expense (electricity, water, etc.)
transport	Transport / commute expenses
food	Food / groceries expense
entertainment	Entertainment / leisure costs
healthcare	Healthcare / medical expenses
savings	Amount saved (or residual) after expenses
investments	Amount invested (if available)

Column	Description
other_expenses	Any miscellaneous expenses

2.Data Quality

- **Missing values / Nulls** — Are there nulls or blanks in any columns?
- **Outliers** — Some incomes or expenses may be extremely large or small; check for unrealistic values.
- **Data type consistency** — e.g. income, expenses should be numeric; gender, occupation categorical.
- **Duplication** — Check for duplicate records / repeated rows.
- **Skewness / distribution** — Many financial variables tend to be skewed (long tail).
- **Data reliability / bias** — Since this is self-reported data or survey-based, there may be biases (recall bias, underreporting, selective sampling).
- **Range checks & validity** — e.g. age should be in a reasonable human range, expenses shouldn't exceed income by massive amounts, etc.

3. Operations Performed

Loading the data (e.g. via `pandas.read_csv`)

Inspecting data types and converting types (e.g. converting certain columns to int, float, or category)

Handling missing values:

Dropping rows or columns with too many missing values

Imputing missing values (mean / median / mode / custom)

Removing duplicates

Outlier detection and treatment:

Using IQR method, z-score method to identify extreme values

Capping or removing outliers

Feature creation / derived metrics:

Computing **disposable income** = `income - total_expenses`

Expense ratios: e.g. `food_expense / total_expenses`, `rent / income`

Category-wise proportions

Exploratory Data Analysis (EDA):

Univariate analysis (histograms, boxplots)

Bivariate analysis (scatter plots, correlation matrices)

Grouped aggregations (by gender, age group, occupation)

Correlation / covariance analysis to see relationships

Visualizations: bar charts, heatmaps, boxplots, pie charts, etc.

4. Key Insights

Income vs Savings

Individuals with higher income tend to have higher absolute savings, but as a percentage of income, savings rate may not increase proportionally. (i.e. rich people may spend more in absolute terms).

Expense Allocation / Spending Patterns

- Many people spend a large fraction of income on **food, rent, or transport**
- The share of discretionary spending (entertainment, travel) is relatively small for lower income groups
- Households with more dependents tend to have lower per capita savings

Effect of number of dependents

As the number of dependents increases, disposable income per dependent drops, and savings reduce. Some analyses explore correlation between dependents and savings.

Gender / occupation differences

There may be differences in spending habits, savings behavior, or expense profiles across gender or occupation categories.

Outliers & extremes

A small number of respondents may report extremely high incomes or expenses, which can distort averages.

Correlation patterns

- Positive correlation: income & savings
- Negative correlation: expenses / liabilities vs savings
- Inter-expense correlations: e.g.

- suggest optimal allocations or cost-saving strategies (e.g. bulk buying, public transport, rent sharing).
- value of saving regularly, even in small amounts, and investing in low-risk instruments.

5. Recommendations

1. Encourage better budgeting & expense awareness

Many individuals may not be aware of how much of their income is going into each category — providing tools or awareness can help reduce unnecessary expenditure.

2. Targeted financial planning for high-expense categories

If food, transport or rent are dominating budgets, financial advisors or apps could suggest optimal allocations or cost-saving strategies (e.g. bulk buying, public transport, rent sharing).

3. Promote savings & investment behavior

For lower and middle income groups, financial literacy programs can emphasize the value of saving regularly, even in small amounts, and investing in low-risk instruments.

4. Tailor advice by demographic segment

Since behavior differs by occupation, dependents, gender, suggestions should be personalized: e.g. for families with many dependents, focus on reducing variable costs; for single professionals, optimizing rent and discretionary spend.

5. Introduce emergency fund cushions

Encourage setting aside an emergency fund (e.g., 3–6 months' expenses) to avoid dipping into savings/investments for unexpected costs.

6. Monitor outliers & anomalies

In practice, watchers of expense data (e.g. in a financial app) should flag extreme expense or income values for further review or validation.

7. Leverage technology / apps

Mobile apps that categorize spending automatically, send alerts on overspending, and forecast savings could help users stay on track.