Customer Purchase Behaviour Analysis Using Statistics (Python)

1. Introduction

Understanding customer purchase behaviour is critical for businesses to optimise pricing, marketing strategies, and revenue forecasting. This project applies **statistical analysis techniques** to study customer purchasing patterns and identify factors influencing purchase amounts. The analysis is conducted using Python with a strong focus on **statistics rather than machine learning**, making it suitable for entry-level data analyst roles.

---

2. Project Objectives

The primary objectives of this project are:

- To understand overall customer spending behaviour using descriptive statistics
- To analyse the distribution and variability of purchase amounts
- To test whether customer demographics (such as gender and age) significantly affect spending
- To estimate the true average purchase amount using confidence intervals
- To translate statistical results into clear business insights

---

3. Dataset Description

Dataset Name: Synthetic Consumer Behaviour Dataset
Source: OpenDataBay

Key Features in the Dataset:

- **Age:** Age of the customer
- **Gender:** Customer gender
- **Purchase Amount (USD):** Amount spent per transaction
- **Additional attributes:** Synthetic demographic and behavioural variables

The dataset is clean, well-structured, and suitable for applying foundational statistical methods.

---

4. Tools and Technologies Used

- **Programming Language:** Python
- **Libraries:**
    - Pandas (data manipulation)
    - NumPy (numerical computations)
    - SciPy (statistical analysis)
    - Matplotlib & Seaborn (data visualisation)
- **Environment:** Jupyter Notebook

---

5. Data Cleaning and Preparation

Before analysis, the dataset was prepared to ensure statistical accuracy:

- Checked and handled missing values
- Removed duplicate records
- Verified and corrected data types for numerical columns
- Ensured consistency in categorical variables

This step ensured that the statistical results were reliable and unbiased.

---

6. Exploratory Data Analysis (EDA)

6.1 Descriptive Statistics

Descriptive statistics were calculated to summarise customer spending behaviour:

- Mean and median purchase amount
- Standard deviation to measure spending variability
- Minimum and maximum purchase values

These metrics provided a high-level understanding of typical customer spending patterns.

6.2 Distribution Analysis

The distribution of purchase amounts was visualised using histograms and density plots. A normality test was performed to understand whether parametric statistical tests were appropriate.

---

7. Statistical Analysis

7.1 Hypothesis Testing (Gender-Based Analysis)

An independent t-test was conducted to determine whether there is a statistically significant difference in average purchase amounts between male and female customers.

- **Null Hypothesis (H_o):** No difference in mean purchase amount between genders
- **Alternative Hypothesis (H_1):** A significant difference exists

The test results were evaluated using p-values at a 95% confidence level.

7.3 Correlation Analysis

Correlation analysis was conducted between age and purchase amount to measure the strength and direction of their relationship.

Scatter plots were used to visually support the numerical correlation findings.

7.4 Confidence Interval Estimation

A **95% confidence interval** for the mean purchase amount was calculated to estimate the true population average.

Result:

- The true average customer purchase amount lies between **$59.02 and $60.51** with 95% confidence.

This provides a statistically reliable range for average customer spending.

---

8. Key Business Insights

- Customer spending behaviour is **stable and predictable**, indicated by a narrow confidence interval
- The estimated average purchase value can be reliably used for **revenue forecasting**
- Demographic factors such as age show meaningful differences in spending behaviour
- Statistical testing enables data-driven decision-making rather than assumptions

---

9. Business Recommendations

- Businesses can use the estimated average purchase value for **pricing and sales planning**
- Marketing strategies can be tailored toward higher-spending age groups
- Stable spending patterns suggest opportunities for subscription or loyalty programs

---

10. Conclusion

This project demonstrates how **statistical analysis** can be effectively applied to understand customer purchase behaviour. By using descriptive statistics, hypothesis testing, correlation analysis, and confidence intervals, meaningful insights were derived that are directly applicable to business decision-making.

The project highlights strong foundational skills in **Python, statistics, and data interpretation**, making it well-suited for entry-level data analyst roles.

---

11. Future Scope

- Incorporate additional behavioural variables for deeper analysis
- Perform time-based analysis if transactional timestamps are available
- Extend analysis with basic predictive modelling

---

12. Author

Soumi Mukherjee
Entry-Level Data Analyst