STATISTICAL BUSINESS ANALYSIS REPORT

Week 7 – Developers Arena Data Science Internship

---

1. Introduction

In modern data-driven organizations, decisions are increasingly supported by statistical evidence rather than intuition alone.
This project applies **core statistical concepts** to real business datasets in order to evaluate assumptions, uncover relationships, and quantify uncertainty.

The analysis focuses on **sales performance** and **customer churn behavior**, using statistical tools such as descriptive statistics, hypothesis testing, correlation analysis, confidence intervals, and regression modeling.

---

2. Business Problem Statement

The key business questions addressed in this analysis are:

- What is the typical sales performance of the business?

- How variable are sales values across transactions?

- Are observed sales patterns statistically significant?

- Do customer behaviors (such as churn) affect revenue?

- Which factors have a measurable impact on sales outcomes?

Answering these questions helps organizations reduce risk and improve strategic planning.

---

3. Data Description

Two datasets were used:

3.1 Sales Dataset (`sales_data.csv`)

- Contains transactional sales data

- Includes product, pricing, quantity, region, and revenue details

- Used for sales trend and performance analysis

3.2 Customer Churn Dataset (`customer_churn.csv`)

- Contains customer behavioral indicators

- Includes churn status and customer attributes

- Used to compare churned vs retained customers

Both datasets were validated for completeness and correctness before analysis.

---

4. Methodology

The statistical methodology followed a structured approach:

4.1 Descriptive Statistics

Key summary measures such as **mean, median, mode, and standard deviation** were calculated to understand central tendency and variability.

4.2 Distribution Analysis

Histograms and density plots were used to visually assess the distribution of sales data and detect skewness or outliers.

4.3 Correlation Analysis

Pearson correlation coefficients were computed to measure the strength and direction of relationships between numerical variables.

4.4 Hypothesis Testing

Multiple hypothesis tests were conducted to evaluate business assumptions using statistical significance testing.

4.5 Confidence Interval Estimation

Confidence intervals were calculated to quantify uncertainty around estimated means.

4.6 Regression Analysis

Linear regression modeling was applied to identify predictive relationships and quantify their impact.

---

5. Hypothesis Testing Framework

For all hypothesis tests, the following framework was applied:

- **Null Hypothesis (H_0):** No significant effect or difference exists

- **Alternative Hypothesis (H_1):** A significant effect or difference exists

- **Significance Level (α):** 0.05

Decisions were made based on **p-values** relative to the chosen significance level.

---

6. Key Statistical Findings

6.1 Descriptive Insights

- Average sales value provides a baseline performance indicator

- Standard deviation highlights variability in transaction values

- Median and mode reveal distribution shape and skewness

6.2 Correlation Results

- Quantity and total sales show strong positive correlation

- Price influences revenue but with varying strength

- Correlation heatmaps visually reinforce numerical findings

6.3 Hypothesis Test Results

- Mean sales differ significantly from benchmark values

- Sales performance varies across customer segments

- Churn status shows measurable impact on revenue patterns

6.4 Confidence Interval Interpretation

The 95% confidence interval indicates the range within which the true mean sales value is likely to lie, providing statistical assurance beyond a single-point estimate.

6.5 Regression Analysis Results

Regression modeling confirms that **quantity sold is a statistically significant predictor of total sales**, with meaningful explanatory power (R^2).

---

7. Business Interpretation

Statistical results translate into actionable insights:

- High variability in sales suggests opportunities for stabilization strategies

- Strong correlations highlight key drivers of revenue

- Churn-related sales differences emphasize the importance of retention programs

- Regression results support data-backed forecasting and planning

---

8. Limitations

- Analysis is limited to available historical data

- External factors such as marketing spend were not included

- Results assume independence of observations

Despite these limitations, conclusions remain statistically sound within scope.

---

9. Conclusion

This project demonstrates how statistical techniques transform raw data into validated business insights.
 By applying hypothesis testing, confidence intervals, and regression analysis, decision-makers gain **quantitative evidence** to support strategic actions.

---

10. Future Scope

- Incorporate marketing and operational datasets

- Apply non-linear regression models

- Perform time-series statistical analysis

- Extend to predictive and prescriptive analytics