**Customer Insights – A Statistical Investigation**

**Project Title:**

Unlocking Customer Insights: A Statistical Investigation

**Activities:**

Apply descriptive statistics, exploratory data analysis, visualizations, and hypothesis testing to investigate behavioural patterns, spending trends, and demographic influences in a real-world customer dataset.

**Business Problem Statement:**

A mid-sized Indian retail company is sitting on a treasure trove of customer data — but lacks clarity on what story it tells. They need a team of analysts to conduct a rigorous statistical investigation to validate business assumptions, uncover hidden patterns, and support data-driven decision-making.

As a data analytics intern, you have been tasked with:

* Using statistical tools to explore and summarize the dataset
* Comparing customer behavior across demographic segments
* Testing business hypotheses using statistical methods

**Key Goals:**

* Analyze customer demographics and behavior using descriptive statistics
* Visualize spending, frequency, and demographic attributes
* Segment customers using appropriate quantitative logic
* Test statistical hypotheses around:
* Age and spending
* Gender and transaction frequency
* Geography and engagement
* Deliver clear, statistically valid insights

**Dataset Description**

You are provided with a synthetic dataset of **10675 rows** for roughly **1,000 unique customers**. The dataset exhibits Behavioural Signals like Spending habits, interaction recency, and lifestyle choices like pet ownership

The dataset description:

| **Column Name** | **Description** |
| --- | --- |
| CustomerID | Unique identifier for each customer |
| Name | Full name of the customer (US names used) |
| State | US state where the customer resides |
| Education | Highest level of education attained (High School to PhD) |
| Gender | Gender identity (Male, Female, Non-Binary) |
| Age | Age of the customer (between 18 and 80) |
| Married | Marital status: Yes or No |
| NumPets | Number of pets owned by the customer (0 to 4) |
| MonthlySpend | Amount spent monthly in USD, generated using a skewed Gamma distribution |
| DaysSinceLastInteraction | Number of days since last interaction with the company |

**Step-by-Step Investigation Plan**

**Note:** Wherever Statistical Computation is applicable, you will need to use Python and associated libraries

Step 1: Understand Your Data

Business Purpose: Familiarize yourself with who your customers are and what attributes are available to you.

* Load and preview the dataset.
* Check data types, unique values, and presence of nulls.
* Understand which variables are categorical and which are numerical.

Step 2: Descriptive Statistics

Business Purpose: Describe your customer base — how old are they, how much do they spend, are they active?

* Compute:
  + Mean, median, std dev for Age, MonthlySpend, DaysSinceLastInteraction
  + Mode for categorical variables: Gender, Education, Married

Step 3: Data Visualization

Business Purpose: Reveal patterns that numbers alone can’t show.

* Plot histograms and boxplots for Age, MonthlySpend
* Create a bar chart for Gender, Education, State
* Scatterplot: Age vs MonthlySpend
* KDE: Spending behavior by education level or marital status

Step 4: Bivariate Analysis

Business Purpose: Check how customer attributes relate to one another.

* Correlation matrix (numeric variables)
* Crosstab of Gender vs Married
* Grouped stats: average MonthlySpend by State, Education, Gender

Step 5: Formulate Hypotheses

Business Purpose: Turn business questions into statistical tests.

| **Business Question** | **Statistical Test** |
| --- | --- |
| Do males and females spend differently? | Independent t-test |
| Does education level impact average monthly spend? | One-way ANOVA |
| Is marital status related to the number of pets owned? [Only for DS Students] | Chi-square test |
| Are older people less active? | Correlation (Age vs DaysSinceLastInteraction) |
| Does state-wise spend vary significantly? | ANOVA |

Step 6: Run Hypothesis Tests

Business Purpose: Validate or reject your assumptions with confidence.

* Define null and alternate hypotheses
* Choose test based on data types
* Check assumptions: normality, independence, homogeneity of variance
* Interpret p-values and confidence intervals

Step 7: Present Business Insights

Business Purpose: Translate stats into strategy.

Create 4–5 takeaways. For example:

* “Customers with Master’s degrees spend 18% more per month on average.”
* “Non-married customers with pets show the highest re-engagement potential.”
* “Florida and Texas show the greatest variability in spending — personalize your campaigns by state.”