INTERNSHIP REPORT FOR TWILEARN

---

1. Introduction

My 2- months internship at TwiLearn as Data Science Intern included one month of online training sessions and further project development. I had the opportunity to work on two significant projects : HR Analytics Dashboard and Zomato Data Analysis.This internship helped me to gain experience in data analytics through practical applications, focusing on data collection, data cleaning & pre-processing , transformation, data analysis and data visualization.

2. About Projects

The problem statement and dataset were provided for both the projects. For the first one a dashboard was to be created to present the valuable insights and the second project included answering certain questions through analysis and visualizations.

A) Project 1: HR Data Analytics

Problem Statement :
To create an interactive dashboard that would enable HR to analyze and further make data-driven decisions regarding the employee count, attrition count, employee performance and engagement.

Dataset :
The dataset consisted of 26 columns and 1500 entries showing employee details like employee id , gender, age group , attrition , department etc.

Key Objectives :
- To understand attrition count with respect to certain factors like age, gender, salary.

- To visualize key HR metrics that help in workforce planning.

Methods used :

1. Data Cleaning:
    - The data was initially scattered, and various inconsistencies like missing salary data and incorrect age values were rectified through Google Sheets preprocessing steps.
    - Created Pivot Tables to understand key metrics like employee count according to gender .
    - Another pivot table showcased age distribution of employees in the company and further visualized using a line chart.

2. Data Transformation:
    - Imported the Data in PowerBI and used PowerBI Query to perform further transformations.
    - The business travel column consisted of 2 unique entries as travelrarely and travel_rarely.
    - Replaced the value as travel_rarely for all the entries having travelrarely.
    - Ensured that all the columns have a dedicated data type.
    - Removed empty rows and filtered the duplicate counts present in the EmpID column.
    - Replaced values and created conditional columns for converting Attrition Data from categorical to numeric.

3. Dashboard Creation:
    - Identified 7 KPI’s they are Count of Employees , Attrition Count , Average Age count of Employees, Average Salary and Average Years at the company.
    - Attrition Rate was one of the important KPI calculated as Sum of attrition count / Sum of Employee Count which gave the rate as 16.1%.
    - Further these key metrics were used to visualize Attrition by Education, Age , Job role and years at company.

B) Project 2: Zomato Data Analysis

Problem Statement:
Given dataset for Zomato's restaurant data to derive customer preferences , restaurant performance and market trends.

Dataset:
The dataset consisted of 7 columns that had restaurant-related information namely name of restaurants , ratings , votes , listed type , online order , table booking and cost for two. It helps to understand customer preferences and restaurant performance.

Key Objectives :
- Analyze Customer Preferences: Identify popular restaurant types and services like online ordering and table booking.
- Evaluate Restaurant Performance: Compare ratings, votes, and costs to assess restaurant performance.
- Identify Market Trends: Explore rating trends and customer spending habits for actionable insights.

Methods Used:

1. Data Cleaning and Preprocessing :
    - Cleaned the ratings columns to extract numerical values from text entries.
    - Ensured no null values were present in the dataset.

2. Exploratory Data Analysis:
   Conducted thorough analysis to answer key business questions like :
    - Identification of most popular restaurant types among customers.
    - Analysis of the total votes received by each restaurant type.
    - Insights into average spending habits of couples ordering food online.
    - Comparison of ratings between online and offline orders.

3. Visualizations :
   Created various visualizations to illustrate findings, including:
   - Count Plots to depict customer preferences for different restaurant types.
   - Line plots to visualize voting trends across restaurant types.
   - Histograms representing the distribution of ratings.
   - Heatmaps to observe online versus offline orders for each restaurant type.

Key Insights:

- Dining is most customer's first preference and has received maximum votes also.
- The majority of ratings is observed to be between 3.5 to 4.
- Online orders have received maximum ratings then offline.
- Dining restaurants observe more offline orders.

3. Skills Gained

- Data Cleaning & Preprocessing: Proficiency in handling missing or inconsistent data and transforming it for analysis.

- Data Analysis & Visualization: Experience with creating data visualizations using tools like Power BI, Seaborn, and Matplotlib to derive insights.

- Dashboard Creation: Learned to build interactive dashboards to display key metrics.

- Problem Solving: Enhanced ability to approach and solve real-world business problems through data analysis.

4.Tools and Technologies :

- Google Sheets: For initial data cleaning and preprocessing.

- Power BI: To perform advanced data analysis and build interactive dashboards.

- Python (pandas, NumPy): For data cleaning and analysis in the Zomato Data Analysis project.

- Seaborn & Matplotlib: To create visualizations for exploratory data analysis.

- Google Colab: For running Python code and sharing analysis results.