Data Analysis Project- Using MRM Predicting Restaurant Revenue

Introduction

In the dynamic and highly competitive restaurant industry, accurate revenue forecasting is a crucial component of strategic planning and operational efficiency. This project focuses on developing a **Multiple Regression Model (MRM)** using MS Excel to predict restaurant revenue by identifying and analyzing key determinants. The model is designed to provide actionable insights that enable restaurant managers to optimize their operations, allocate resources effectively, and enhance profitability.

Key operational and financial variables, such as **seating capacity**, **marketing budget**, **chef experience**, and other factors, were analyzed to uncover their impact on revenue. The project adopted a structured approach, starting with rigorous data cleaning and exploratory data analysis (EDA) to ensure data quality and integrity. Using statistical modeling, the project achieved a **Multiple R of 0.966** and an **R-Square of 0.933**, signifying that the model explains **93.3%** of the variance in revenue. This high level of accuracy underscores the model's reliability and practical utility.

The insights derived from this analysis are instrumental in identifying revenue drivers, such as high-performing cuisine types or strategic marketing investments, that significantly influence profitability. By leveraging these findings, the project provides strategic recommendations to improve financial risk assessment and support data-driven decision-making, ultimately paving the way for sustainable growth in the restaurant business.

OBJECTIVE

The objective of this project is to develop a reliable and accurate **Multiple Regression Model (MRM)** to predict restaurant revenue and identify key factors driving profitability. By leveraging MS Excel for data analysis and modelling, the project aims to:

1. Predict Restaurant Revenue: Accurately forecast revenue using key operational and financial variables such as seating capacity, marketing budget, and chef experience.

2. Identify Revenue Drivers: Quantify the impact of critical factors like cuisine type and location on revenue generation to prioritize high-performing areas.

3. Enable Strategic Decision-Making: Provide actionable insights for optimizing resource allocation, marketing strategies, and operational efficiency.

4. Enhance Financial Risk Assessment: Support planning by identifying risks and opportunities for improving revenue stability.

5. Improve Operational Efficiency: Offer practical recommendations for boosting profitability through data-driven adjustments in operations.

Steps Involved in Predicting Restaurant Revenue

Data Collection and Preprocessing – Removing Outliers Using the IQR Method

In the first step of this project, we focus on preprocessing the data by identifying and removing outliers. Outliers are extreme values that deviate significantly from the rest of the data and can skew the results of statistical models. In this case, we are working specifically with the **Marketing Budget** variable and will use the **Interquartile Range (IQR)** method to detect and remove outliers.

Process to Remove Outliers Using the IQR Method:

1. **Calculate the Quartiles and IQR:**
    - **Quartile 1 (Q1):** 1889
    - **Quartile 3 (Q3):** 4008.5
    - **Interquartile Range (IQR):** Q3 - Q1 = 4008.5 - 1889 = **2119.5**
2. **Determine the Outlier Boundaries:** The outlier boundaries are calculated using the formula:
    - **Lower Bound:** Q1 - 1.5 * IQR = 1889 - 1.5 * 2119.5 = **-1290.25**
    - **Upper Bound:** Q3 + 1.5 * IQR = 4008.5 + 1.5 * 2119.5 = **7187.75**
3. **Identify Outliers:**
    - Any data point in the **Marketing Budget** column that is below **-1290.25** or above **7187.75** is considered an outlier.

4. **Create a New Column to Indicate Outliers:** To manage the outliers, we add a new column that labels each data point as either **"Keep"** or **"Remove"** based on whether it lies within the bounds or not.
    - **If the value is between -1290.25 and 7187.75** (i.e., within the boundaries), the entry will be labeled as **"Keep"**.
    - **If the value is less than -1290.25 or greater than 7187.75**, the entry will be labeled as **"Remove"**.
5. **Remove Outliers:** After marking outliers, you can filter the data to exclude all the rows where the **"Remove"** label appears, thereby ensuring that only the valid data remains for further analysis.

Practical Example in Excel:

- Let's say the marketing budget for one of the restaurants is **-1300**, which falls below the lower bound of **-1290.25**. This value will be identified as an outlier and labeled as **"Remove"** in the new column.
- Another restaurant has a marketing budget of **7200**, which is above the upper bound of **7187.75**, so it will also be marked as **"Remove"**.

By removing the outliers, we ensure that the data used for further analysis (like regression modeling) is not biased or distorted by extreme values. This process is crucial because the presence of outliers can lead to inaccurate model predictions and unreliable insights.
After identifying and removing outliers from the **Marketing Budget** data using the IQR method, we have a clean dataset where only valid data points are retained for analysis. This step helps ensure the accuracy and reliability of the subsequent steps in the project, particularly in predicting restaurant revenue using regression models.

Graphical Analysis

Graph 1: Revenue vs. Seating Capacity

- **Description**: This graph depicts the relationship between seating capacity and the revenue generated by restaurants.
- **Insights**:
    - **Trend**: There is a positive linear relationship between seating capacity and revenue, as indicated by the increasing trend line.
    - **Interpretation**: Restaurants with higher seating capacities tend to generate more revenue, which makes sense as they can serve more customers simultaneously.
    - **Actionable Insight**: To increase revenue, investing in larger seating capacities could be a viable strategy for restaurants, provided demand justifies the expansion.

Graph 2: Revenue vs. Cuisine

- **Description**: This bar chart shows the revenue generated by different cuisines offered by the restaurant.
- **Insights**:
    - **Highest Revenue**: Japanese cuisine generates the highest revenue.
    - **Second Place**: French cuisine comes next, followed by Italian cuisine.
    - **Lowest Revenue**: Mexican cuisine generates the least revenue.
    - **Interpretation**: Customer preferences and pricing might be driving the differences in revenue among cuisines.
    - **Actionable Insight**: Focus on popular cuisines like Japanese and French to maximize profitability. Consider revising strategies for low-performing cuisines like Mexican.

Graph 3: Location-Wise Revenue

- **Description**: This chart highlights the revenue generated by restaurants located in different areas: Downtown, Suburban, and Rural.
- **Insights**:
    - **Downtown Revenue**: Downtown locations generate the highest revenue.
    - **Suburban Revenue**: Suburban locations come second in revenue generation.

- **Rural Revenue**: Rural areas contribute the least to revenue.
- **Interpretation**: Downtown areas likely have higher foot traffic and customer density, leading to higher revenues.
- **Actionable Insight**: Invest more in Downtown areas and optimize marketing efforts in Suburban regions to boost performance.

Graph 4: Average Revenue by Cuisine

- **Description**: This bar chart represents the average revenue generated by each cuisine.
- **Insights**:
  - **Top Performers**: Japanese and French cuisines lead in average revenue generation.
  - **Mid Performers**: Italian cuisine performs moderately well.
  - **Low Performers**: Indian and Mexican cuisines have the lowest average revenue.
  - **Interpretation**: High-performing cuisines may be priced higher or more popular among customers.
  - **Actionable Insight**: Focus on promoting and diversifying menu offerings for popular cuisines while reassessing pricing or marketing strategies for lower-performing ones.

Correlation Values

1. **Correlation Between Marketing Budget and Revenue**: 0.33
   - **Explanation**: A weak positive correlation suggests that while marketing has some impact on revenue, it is not the sole factor driving it.
   - **Actionable Insight**: Focus on optimizing marketing strategies rather than just increasing the budget.
2. **Correlation Between Seating Capacity and Revenue**: 0.68
   - **Explanation**: A strong positive correlation indicates that seating capacity significantly influences revenue.
   - **Actionable Insight**: Expanding seating capacity could be an effective strategy to increase revenue.

Recommendations

1. **Optimize Seating Capacity**: Expand seating in high-demand locations, especially Downtown, to capitalize on the strong correlation with revenue.
2. **Focus on Popular Cuisines**: Prioritize Japanese, French, and Italian cuisines for promotions and expansion.
3. **Improve Performance in Rural Locations**: Evaluate customer preferences and adjust strategies to improve performance in rural areas.
4. **Optimize Marketing Strategies**: Since marketing has a weak correlation with revenue, focus on targeted campaigns to achieve better returns on investment.

Descriptive Analysis

The following section provides an in-depth analysis of the descriptive statistics for the dataset, focusing on three key variables: **Average Meal Price**, **Marketing Budget**, and **Revenue**. These insights help us understand the central tendencies, variations, and distributions within the data.

---

1. Average Meal Price

This variable represents the price of meals offered across various locations.

- **Mean**: The average meal price is **47.88**, indicating that, on average, meals are priced around this value.
- **Median**: The median price is **45.5**, suggesting that half of the meal prices are below and half are above this value. The proximity of the median to the mean suggests minimal skewness in the distribution.
- **Mode**: The most frequently occurring meal price is **59.98**, possibly reflecting a popular pricing strategy or customer preference.

- **Standard Deviation**: The standard deviation of **14.35** reflects moderate variability in meal prices, meaning most prices are within ±14.35 units of the mean.
- **Range**: The range of **51** highlights a significant difference between the minimum price (**25**) and maximum price (**76**), demonstrating the diversity in pricing strategies.
- **Kurtosis**: A kurtosis value of **-1.15** suggests a flat distribution (platykurtic), indicating fewer extreme values in meal prices.
- **Skewness**: A positive skewness of **0.31** indicates a slight right-skewed distribution, implying the presence of a few higher-priced meals.
- **Sum and Count**: The total sum of all meal prices is **382,030.07**, with **7,979** data entries.

Interpretation: The distribution of meal prices is relatively balanced, with minor skewness and moderate variability. The mode at **59.98** suggests this could be an optimal price point to attract customers.

---

2. Marketing Budget

This variable captures the amount allocated for marketing by various restaurants.

- **Mean**: The average marketing budget is **2,977.47**, providing a benchmark for marketing expenditure.
- **Median**: The median marketing budget is **2,754**, indicating that half the marketing budgets are below and half are above this value. The mean being slightly higher than the median suggests a right-skewed distribution.
- **Mode**: The most common marketing budget amount is **3,388**, which may represent a standard marketing spend across several locations.
- **Standard Deviation**: The standard deviation of **1,491.17** indicates considerable variability in marketing budgets, showing differences in spending across locations.
- **Range**: A range of **6,578** reveals a large disparity between the minimum (**604**) and maximum (**7,182**) marketing budgets.
- **Kurtosis**: A kurtosis value of **0.08** indicates a near-normal distribution with moderate tails.
- **Skewness**: A skewness of **0.79** suggests a moderately right-skewed distribution, implying some restaurants invest significantly more in marketing.
- **Sum and Count**: The total marketing budget across all entries is **23,752,727**, with **7,979** data entries.

Interpretation: The marketing budget shows significant variation, with some locations investing much higher amounts. The positive skewness suggests the need for more standardized budget allocations.

---

3. Revenue

This variable reflects the revenue generated by restaurants across locations and cuisines.

- **Mean**: The average revenue generated is **645,180.8**, indicating the central tendency of revenue performance.
- **Median**: The median revenue is **592,910.7**, slightly below the mean, indicating a right-skewed distribution.
- **Mode**: No mode exists for revenue, suggesting unique revenue values for most locations.
- **Standard Deviation**: A standard deviation of **263,282.2** reveals significant variability in revenue, indicating differing performance levels across restaurants.
- **Range**: The revenue range of **1,341,759** (minimum **604** to maximum **1,534,186**) demonstrates the vast differences in performance.
- **Kurtosis**: A kurtosis value of **0.11** suggests a slightly peaked distribution compared to normal.
- **Skewness**: The skewness of **0.77** indicates a moderately right-skewed distribution, meaning a few restaurants generate much higher revenue.
- **Sum and Count**: The total revenue is **5.15 billion**, with **7,979** entries.

Interpretation: Revenue data shows a large variation, with some restaurants significantly outperforming others. This variability is likely influenced by factors such as location, cuisine, and marketing effectiveness.

Summary of Insights

1. **Meal Pricing**: Meal prices show moderate variability, with most prices clustered around the mean. The mode price (**59.98**) might indicate an optimal pricing strategy.

2. **Marketing Budgets**: Marketing budgets are highly varied, with a moderate positive skew. Locations with higher budgets might require further analysis to understand the return on investment.
3. **Revenue**: Revenue data reflects significant differences in performance across restaurants. The right skew suggests the need to analyze top-performing restaurants for best practices.

Actionable Recommendations:

1. Standardize pricing strategies to align closer to the popular price point (**59.98**) while accommodating variations for specific cuisines.
2. Evaluate marketing budget allocations to ensure efficiency and maximize ROI, focusing on standardizing budgets across similar-performing locations.
3. Perform deeper analysis on top revenue-generating restaurants to replicate their strategies in underperforming areas.

Here's a detailed explanation structured for your project document. It's designed to be clear and comprehensive, fitting a professional project format.

Analysis of Multiple Regression Results

1. Overview of Regression Model

The multiple regression model was developed to analyze the factors influencing restaurant revenue. This section presents the results of the regression analysis, including the significance of predictors, the goodness-of-fit metrics, and actionable insights derived from the model.

---

2. Model Evaluation Metrics

The key evaluation metrics provide insight into the model's performance:

- **Multiple R (0.966):** This value indicates a strong positive linear relationship between the predictors (independent variables) and the target variable (restaurant revenue). A value close to 1 reflects a high degree of fit.
- **R Square (0.933):** This metric signifies that 93.3% of the variability in restaurant revenue is explained by the predictors in the model. It suggests a very high explanatory power of the model.
- **Adjusted R Square (0.932):** Adjusted R Square takes into account the number of predictors in the model and prevents overestimation of the model's fit. This value, being close to R Square, confirms the reliability of the model.
- **Standard Error (69,185.89):** The standard error measures the average deviation of the predicted restaurant revenue from the actual values. A lower value of this metric indicates a better fit.

---

3. Analysis of Variance (ANOVA)

The ANOVA table helps in understanding whether the overall regression model is statistically significant:

- **Regression Sum of Squares (5.58337E+14):** This indicates the variability in revenue that is explained by the predictors.
- **Residual Sum of Squares (3.9988E+13):** Represents the unexplained variability in the revenue.
- **F-statistic (8,972.596):** A very high F-statistic value demonstrates the model's overall strength.
- **Significance F (0):** The p-value for the F-test is 0, which indicates that the model is statistically significant, and at least one of the predictors contributes to explaining the variability in revenue.

---

4. Coefficient Interpretation

Each predictor's coefficient represents its effect on restaurant revenue, holding all other variables constant.

Key Predictors

1. **Seating Capacity (10,706.51):**
    - Revenue increases by approximately $10,706 for every additional unit of seating capacity.

- **Significance:** The p-value (0) indicates this is a highly significant predictor, making seating capacity a crucial determinant of revenue.

2. **Marketing Budget (0.402):**
   - This predictor shows a negligible increase in revenue per dollar spent on marketing.
   - **Significance:** A high p-value (0.88) indicates that marketing budget is statistically insignificant in this model.
3. **Social Media Followers (0.19):**
   - While having more social media followers shows a slight positive effect on revenue, it is statistically insignificant (**p-value = 0.467**).
4. **Chef Experience Years (876.44):**
   - Revenue increases by $876 for each additional year of chef experience.
   - **Significance:** With a p-value < 0.001, this variable is statistically significant, emphasizing the importance of experienced chefs.
5. **Ambience Score (59.50) & Service Quality Score (244.71):**
   - Both variables exhibit minimal impact on revenue and are statistically insignificant (**p-values > 0.4**).

Cuisine Types

Cuisine types significantly influence revenue, as indicated by their coefficients:

- **Japanese Cuisine (196,548.66):**
  - The most impactful cuisine type, leading to a substantial increase in revenue compared to the baseline (American cuisine).
  - **Significance:** Highly significant (**p-value = 0**).
- **Mexican Cuisine (-44,363.74):**
  - A negative impact on revenue, which is statistically significant (**p-value = 0**).
- **French Cuisine (44,433.50):**
  - A positive impact on revenue, also statistically significant (**p-value = 0**).
- **Indian Cuisine (-12,968.53):**
  - This cuisine type negatively affects revenue, and the effect is significant (**p-value < 0.001**).

Location Types

The location of the restaurant also influences revenue:

- **Suburban (2,171.43):**
  - Revenue increases slightly for suburban locations compared to downtown, but the effect is statistically insignificant.
- **Rural (1,505.50):**
  - A marginal increase in revenue compared to downtown, with statistical insignificance.

---

5. Discussion and Insights

From the regression analysis, several important insights emerge:

1. **Key Drivers of Revenue:**
   - Seating Capacity is the most significant predictor, suggesting that optimizing seating arrangements can maximize revenue.
   - Japanese and French cuisines have the highest positive impact on revenue, indicating the popularity of these offerings.
   - Chef Experience Years also positively influences revenue, highlighting the importance of skilled chefs in the restaurant's success.
2. **Underperforming Predictors:**
   - Marketing Budget and Social Media Followers are statistically insignificant, suggesting these areas need re-evaluation to improve their effectiveness.
3. **Negative Impacts:**

- Mexican and Indian cuisines negatively affect revenue. Restaurants offering these cuisines may need to reconsider their pricing, marketing, or menu strategies.

---

6. Recommendations

Based on the analysis, the following recommendations are proposed for restaurant management:

1. **Expand Seating Capacity:**
   - Prioritize increasing seating arrangements where feasible, as this has the largest positive impact on revenue.
2. **Focus on Japanese and French Cuisines:**
   - Restaurants should expand their offerings in these cuisines to capitalize on their strong positive influence on revenue.
3. **Invest in Chef Training:**
   - Providing additional training and retaining experienced chefs can contribute to improved revenue outcomes.
4. **Reassess Marketing Strategies:**
   - Given the minimal impact of marketing budget and social media followers, it is recommended to analyze these channels for better ROI.
5. **Improve Ambience and Service:**
   - Though these factors currently show a negligible impact, improving customer experience could indirectly enhance revenue.
