Data Analysis Portfolio: SQL & RStudio Case Study on DVD Rental Company

Project Background

I worked with a real-world relational database simulating a DVD rental company. The database contained customer transactions, film inventory, payment history, rental dates, staff assignments, and more. I performed a comprehensive analysis using SQL for operational insights and RStudio for statistical exploration, aiming to uncover patterns and provide actionable business recommendations.

Objective

- Operational Insights: Understand customer behavior, film popularity, revenue trends, and staff performance using SQL.

- Statistical Exploration: Perform deeper analysis in RStudio, including customer segmentation, time pattern analysis, and revenue sensitivity assessments.

Skills & Tools

- SQL: PostgreSQL

- R: dplyr, ggplot2, readr

- Techniques: Statistical Testing (t-test, ANOVA), Data Visualization, Business Insights, Sensitivity Analysis

Key Analyses and Insights

1. Customer Revenue Distribution

Business Question: Is there a significant difference in revenue between the top 10% of customers and the rest?

Analysis: Segmented customers into top 10% and bottom 90% based on total revenue contribution. Conducted t-test to evaluate the difference in mean revenue between the two groups.

Result:

- Top 10% mean revenue: ~$151
- Bottom 90% mean revenue: ~$97
- p-value < 2.2e-16, statistically significant

Business Insight:

- The top 10% of customers generate a disproportionately large share of the revenue.

Recommendations:

- Develop a VIP program to retain high-value customers.
- Implement personalized marketing campaigns.
- Conduct A/B testing for promotions targeting high vs. low spending customers.

Sensitivity Analysis:

- Revenue difference becomes insignificant if the top customers reduce spending by more than 34%.

Strategic Insight:

- Business is highly reliant on its top customers, justifying investments in retention and risk mitigation.

2. Daily Revenue Trends

Business Question: What is the pattern of revenue generation throughout the week?

Analysis:

- Aggregated daily revenue.
- Visualized revenue by day of the week.

Initial Findings:

- Mondays and Fridays showed high revenue, but further analysis revealed Monday's spike was an outlier.

Final Findings (after outlier removal):

- Sundays consistently generate the highest revenue.
- Midweek days (Tuesday, Wednesday) underperform.

Recommendations:

- Focus marketing efforts on midweek to boost revenue.
- Optimize staff scheduling for peak days (Sunday and Friday).

3. Rental Duration by Movie Category

Business Question: Are some movie categories rented for longer durations than others?

Analysis:

- Conducted ANOVA to test differences in rental durations by category.

Result:

- Statistically significant differences exist ($p = 0.0469$).
- No specific pairwise differences were significant (Tukey's test).

Business Insight:

- While there is variation in rental duration across categories, differences are minor and should not influence pricing strategies.

Key Takeaways

- Identified high-value customer segments critical to revenue.
- Provided strategic recommendations for customer retention and revenue optimization.
- Performed sensitivity analysis to assess business risks.
- Visualized and interpreted daily revenue trends to optimize operational strategies.
- Enhanced skills in SQL querying, statistical testing, and data visualization.

Full Code and Analysis

For a detailed view of the SQL queries and R scripts, please visit my GitHub repository:
[GitHub Repository - SQL & R Data Analysis]

Let's Connect!

If you're looking for a detail-oriented, analytical, and business-savvy Data Analyst, feel free to connect with me on LinkedIn!