The Rain Effect: A Correlation Analysis of Precipitation, Visit Volume, and Spending

Tools: Python, SQL (PostgreSQL), Power BI, Google Sheets

The Business Question: How do rainy evenings in August influence customer behavior, specifically regarding arrival timing, visit volume, and average spend?

Key Findings Analysis of the **POS data** revealed that rain **does not** significantly alter arrival times or individual spending habits. However, it triggers an increase in both **visit volume**, resulting in a **"Floor Compression" effect** (higher customer density requiring optimized floor management)

To protect client confidentiality, all data used in this analysis has been anonymized, and any names or identifying information related to the business have been removed or modified.

I was hired as a freelance data analyst by **North Dinner**, a small restaurant located in a touristic town in Quebec, to conduct an analysis on how rainy evenings affect restaurant operations. The objective of this project is to provide **data-driven insights** through analysis and visualization that could inform staffing and preparation decisions. This case study follows the **six phases of the data analysis process**.

Business task

The primary objective of this case study is to analyze historical weather data alongside restaurant POS data to better understand how rainy evenings affect North Dinner's operations. By connecting these two data sources, the analysis aims to identify patterns in customer behavior and demand. The resulting insights are intended to support the management team in planning future operations, particularly by improving staffing decisions and preparation processes. While the findings are exploratory in nature, they may help North Dinner reduce operational inefficiencies and better allocate resources during peak periods.

Stakeholders

Sam, the General Manager, is responsible for overseeing restaurant operations. He provided the POS data used in this analysis and is the primary stakeholder who will receive and use the insights generated from this case study.

Prepare process

This analysis is based on two data sources. Historical weather data was obtained from **Open meteo**, a third-party provider of historical weather information. Restaurant POS data was provided directly by the stakeholder and contains transaction-level information used to analyze customer behavior and operational performance.

Data cleaning and preparation

Data Engineering Methodology To overcome the limitations of the legacy Veloce POS system (proprietary `.pry` export format), I engineered a **Python-based ETL pipeline**. This process parsed converted PDF reports, validated the data against control totals to ensure integrity, and loaded the dataset into a structured **PostgreSQL** relational database.

For a detailed breakdown of the PDF parsing script and database schema, please refer to the ***[Data Cleaning Documentation]****.*

The final dataset includes two full years of daily POS data, covering **August 2024 and August 2025**, which form the basis of this analysis.

I queried the necessary data for both August 2024 and 2025 directly from the database. The resulting tables were then cleaned in **Google Sheets**, where I removed any $0 bills and duplicate records to ensure data accuracy before analysis.

For the weather data, I downloaded monthly reports in CSV format from **Open meteo**. There were no missing values for the period under analysis. The CSV files were then converted into **Google Sheets** to ensure compatibility with my PostgreSQL database. To verify data integrity, I cross-checked key values against **Weather Spark** reports for the same dates.

I then created a temporary table in **PostgreSQL** to store the weather data, which allowed me to efficiently query and extract the specific information needed for the analysis.

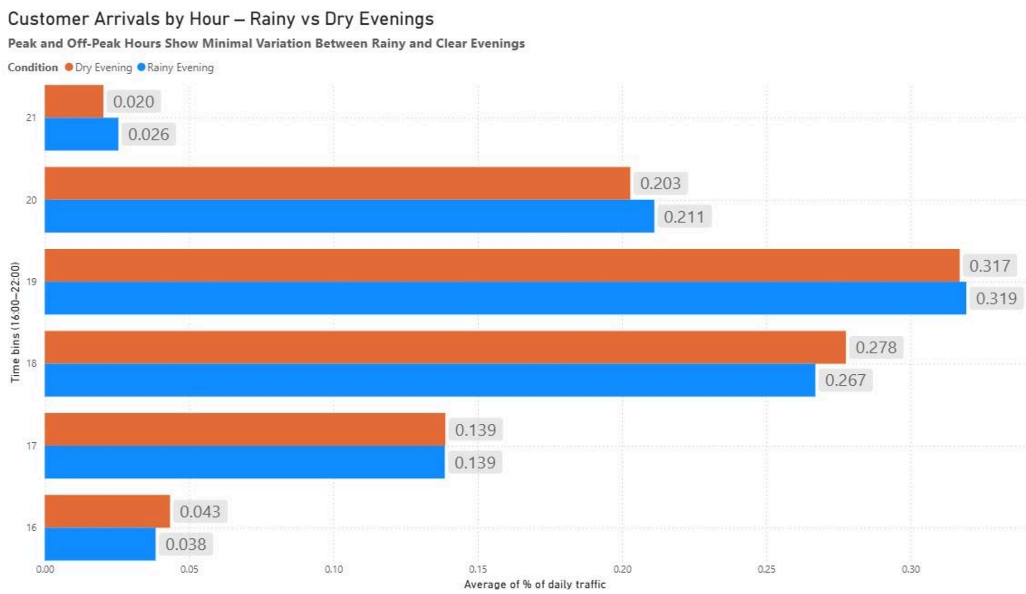
Methodology

The analysis combined historical weather data with restaurant POS data to evaluate how rainy evenings affect operations. Daily customer volume and bill metrics were aggregated from the POS database, with extreme values removed using the 1.5 × IQR rule to minimize outlier influence. Weekly baselines were calculated for each metric, and **daily deviations** were computed as the percentage difference from the weekly average. Rainy and dry days were then compared across these deviations to assess patterns in customer timing, volume, and spending. All data processing, cleaning, and calculations were performed using Python, Google Sheets, and PostgreSQL. Detailed formulas, scripts, and database queries are included in the appendix.

Analysis

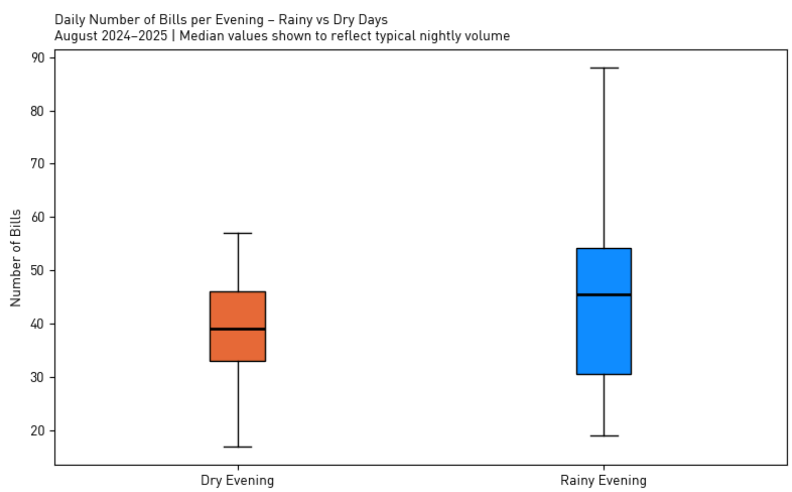
1.Customer timing:

Rain does not appear to have a major effect on the times customers arrive at the restaurant. The difference in volume between the hours is less than 1% between sunny and rainy days. These differences are not considered significant.



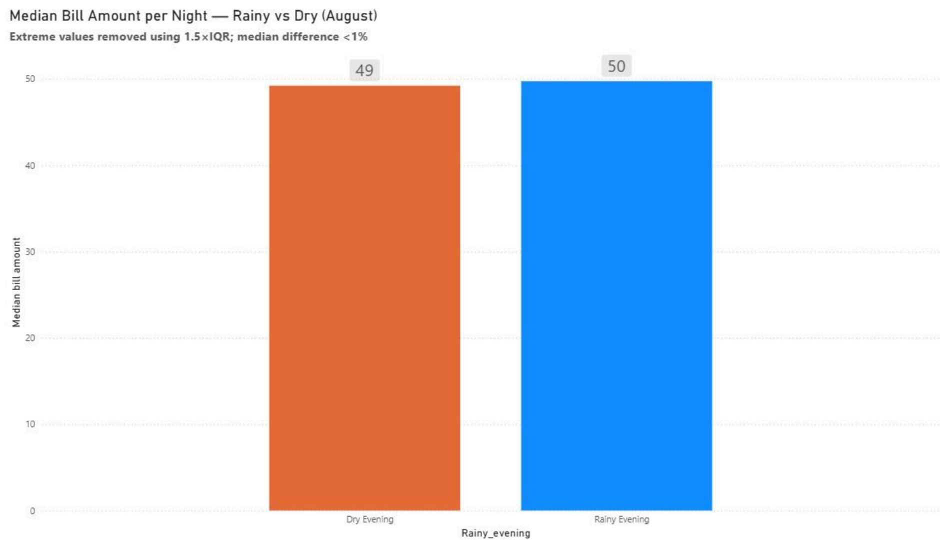
2.Average number of bills per night:

Rain is a high-variance event. While 80% of rainy days track normally, the remaining 20% create massive volume spikes (+14% median), necessitating a 'Surge Protocol' for staff.



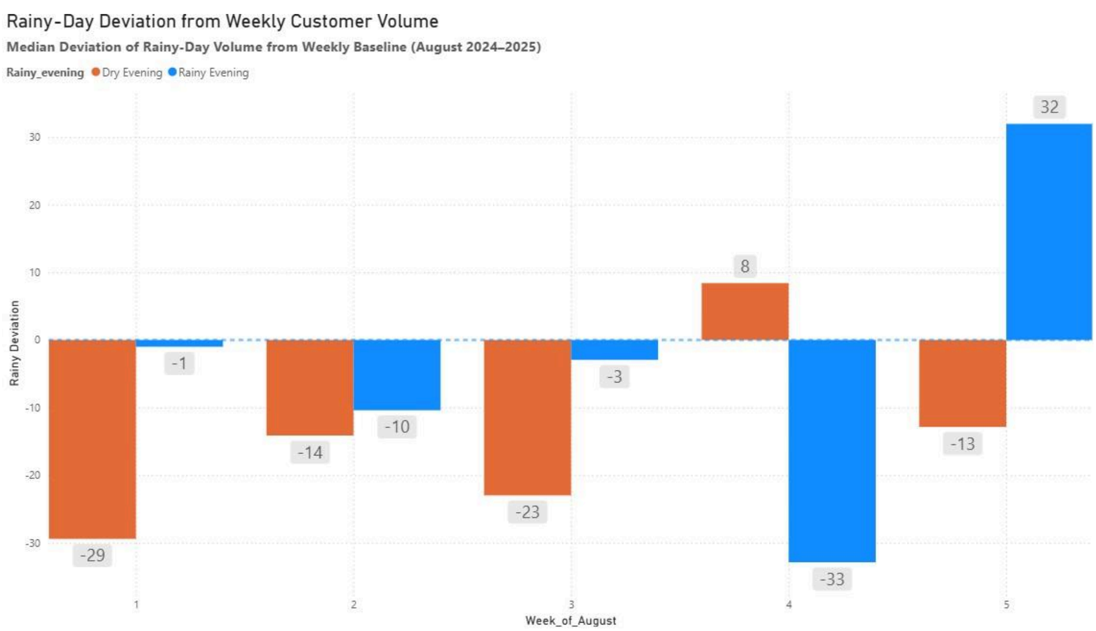
3.Average bill amount per night:

Average spending per customer shows a small increase (+2%) on rainy days. However, due to the small sample size, this change is not statistically significant. After removing extreme bill amounts using the standard 1.5 × IQR method, the median difference between rainy and dry days is less than 1%, indicating minimal effect of rain on individual spending.



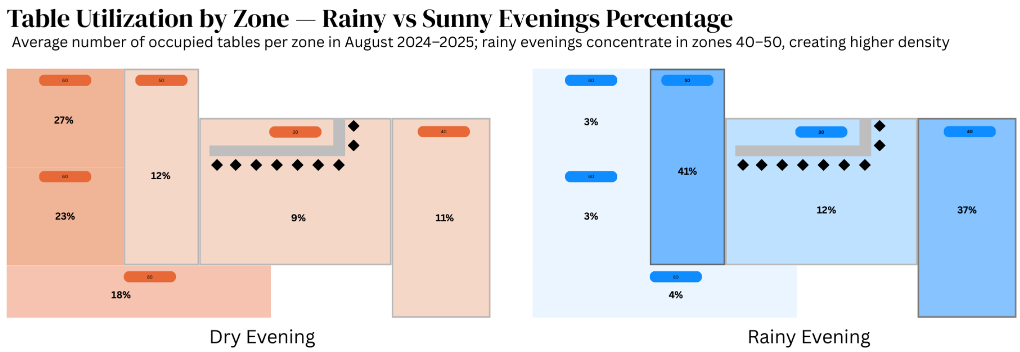
4.Week-by-week deviations:

Across August weeks, the median deviation of rainy days is below the weekly baseline in 4 out of 5 weeks. This indicates that a typical rainy day does not consistently increase volume. However, the mean deviation pooled across all weeks is positive, suggesting that occasional high-volume rainy days raise the overall average effect.



5.Floor compression :

There is a significant shift in table utilization during rainy periods. While sunny days show a balanced distribution across all zones (30s through 80s), rainy days show a concentration in the 40s and 50s.



This suggests that rain effectively "closes" certain sections of the restaurant (likely patio)

This concentration creates "artificial peaks." Even if total volume only increases by 5%, the **density** in the utilized sections increases much more sharply, potentially straining staff in those specific zones.

Share and act

Based on the analysis, it appears that rain does influence customer behavior at North Dinner during August. The restaurant seems to act as a refuge on rainy evenings, which is supported by the slight increase in average bills. Additionally, customers appear to concentrate in certain areas of the restaurant, likely avoiding the outdoor section.

Despite the higher number of bills on rainy nights, customers generally arrive at the same times, and the average bill per customer does not increase significantly. This suggests that peak hours may become slightly busier, but individual spending remains largely unchanged.

Recommendation :

1. Do **not** change opening hours or shift start times based on rain forecasts. Instead, maintain standard staffing levels during peak dinner hours, as rain does not materially affect when customers arrive.

2. On rainy evenings, modestly increase **front-of-house readiness** (e.g., one additional server on standby or faster table turnover processes), especially during peak hours.

Support staff can be increased to manage the door, help the bar and to run the food.

3. **Prioritize compact bookings:** Hosts should give priority to smaller parties to ensure faster table turnover and minimize potential capacity constraints on busy or rainy evenings.

4. **Optimize coverage of the “dead zone”:** If possible, make the underused area of the restaurant more weatherproof, ensuring it can accommodate additional customers during rainy days and handle the slight increase in activity volume. So can the 11% increase in bill volume be better captured.

Executive Summary

Analysis of August 2024–2025 data reveals rain **can** significantly increase volume (up to +14%), but it is **volatile**. It doesn't guarantee a busy night, but it creates the *risk* of a massive surge that overrides the usual trends. This creates "Floor Compression," where interior sections become overcrowded while patio-adjacent zones sit empty. Operations should focus on reallocating staff to interior zones rather than changing shift times.