

Analysis Report

CloudWalk Monitoring Test

This report is a fragment of my solution to the CloudWalk Monitoring Test. This test was taken as part of a selection process to a position on the Monitoring team. The challenge consists of two tasks. The first one is an analysis of hypothetical checkout data, and the second is the implementation of a real-time transaction data monitor.

Task 1 - Checkout Data Analysis

Introduction

In this section, we analyze hypothetical checkout data to identify any anomalies in the sales behavior. The data includes sales per hour for today, yesterday, and previous days.

Data Preparation

The data was extracted from CSV files and converted into a suitable format for analysis. We used the following steps:

1. Data Extraction
 - Extract data from CSV files.
2. Data Preparation
 - Convert time to 'HH:MM' format.
3. Data Load
 - Load data into SQLite tables.
4. Data Merging
 - Merge data using SQL queries.

SQL Query

We used the following SQL query to join and analyze the data from the two CSV files:

```
SELECT
  c1.time,
  c2.today,
  c2.yesterday,
  c1.yesterday AS '2d_ago',
  c2.same_day_last_week AS '7d_ago',
  c1.same_day_last_week AS '8d_ago'
FROM
  checkout_1 c1
JOIN checkout_2 c2 ON c1.time = c2.time AND c1.today = c2.yesterday;
```

We noticed that the sales data for 'checkout_2.yesterday' and 'checkout_1.today' matches 100%, indicating that they are identical. This leads us to assume that these reports are from the same day. The joined data provided insights into sales behavior across different days and helped identify the anomaly.

Data Analysis

The sales data was visualized to identify patterns and anomalies. The chart below shows sales by hour for today, yesterday, and previous days.

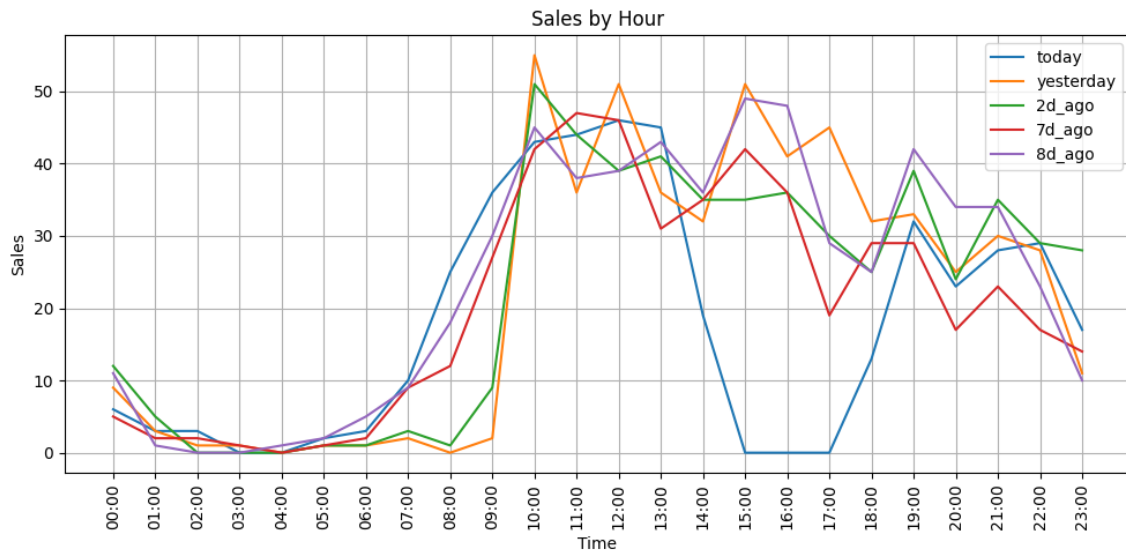


Figure 1: Sales by Hour chart

Pattern and Anomaly Detection

From the chart plotted with the SQL query data, we observed the following pattern:

- Consistent sales patterns during other hours, with peaks around 10:00 and 11:00.

We have also observed the following anomaly:

- Significant drop in sales around 14:00 and 15:00 for today's data, which is not observed on other days.

Conclusion

The analysis revealed a significant drop in sales around 14:00 and 15:00 for today's data, which suggests a potential issue or anomaly. Further investigation is recommended to understand the cause of this drop and take necessary actions.

Based on the sales data, it is plausible that these POS records could come from a business that operates around the clock, such as a 24-hour convenience store like AM-PM or a cafe located at an airport. The consistent sales patterns throughout the day and night support this hypothesis.

Task 2 - Transaction Status Analysis

Introduction

In this section, we analyse the provided transaction data to support the implementation of a monitoring system for real-time transaction data to alert anomalies on transactions status count.

Data Preparation

The data was extracted from CSV files and converted into a suitable format for analysis. The following steps were taken:

1. Data Extraction
 - Extract data from CSV files.
2. Data Preparation
 - Standardize column labels.
 - Merge dataframes.
 - Convert time format to 'HH:MM'.
 - Group by time and status to calculate the average counts.
 - Pivot the data to have statuses as columns.
 - Fill in missing minutes in the dataset.
 - Calculate the percentage of each status per hour.

Data Analysis

The transaction data was visualized to identify patterns and anomalies. The following charts show the average transaction status count over time and the percentage over time of each status:

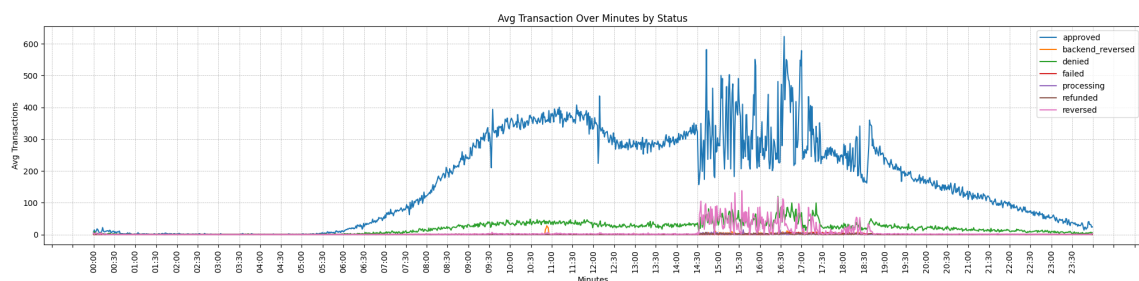


Figure 2: Average Transaction Over Minutes by Status

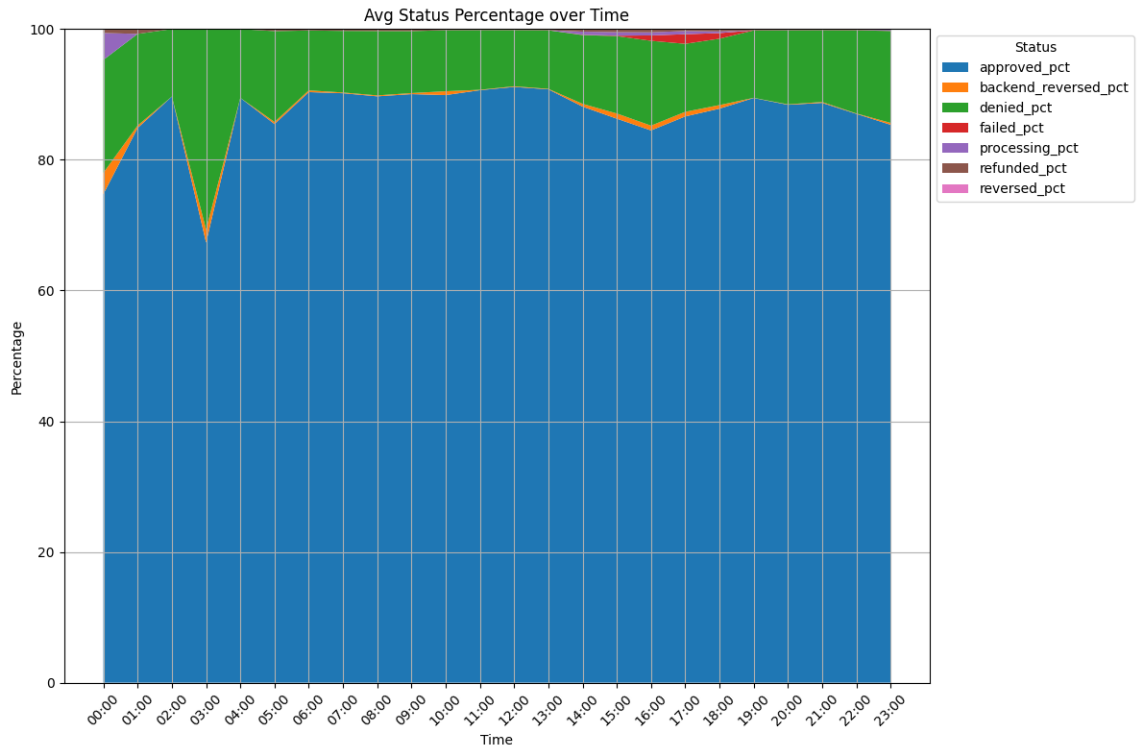


Figure 3: Average Status Percentage over Time

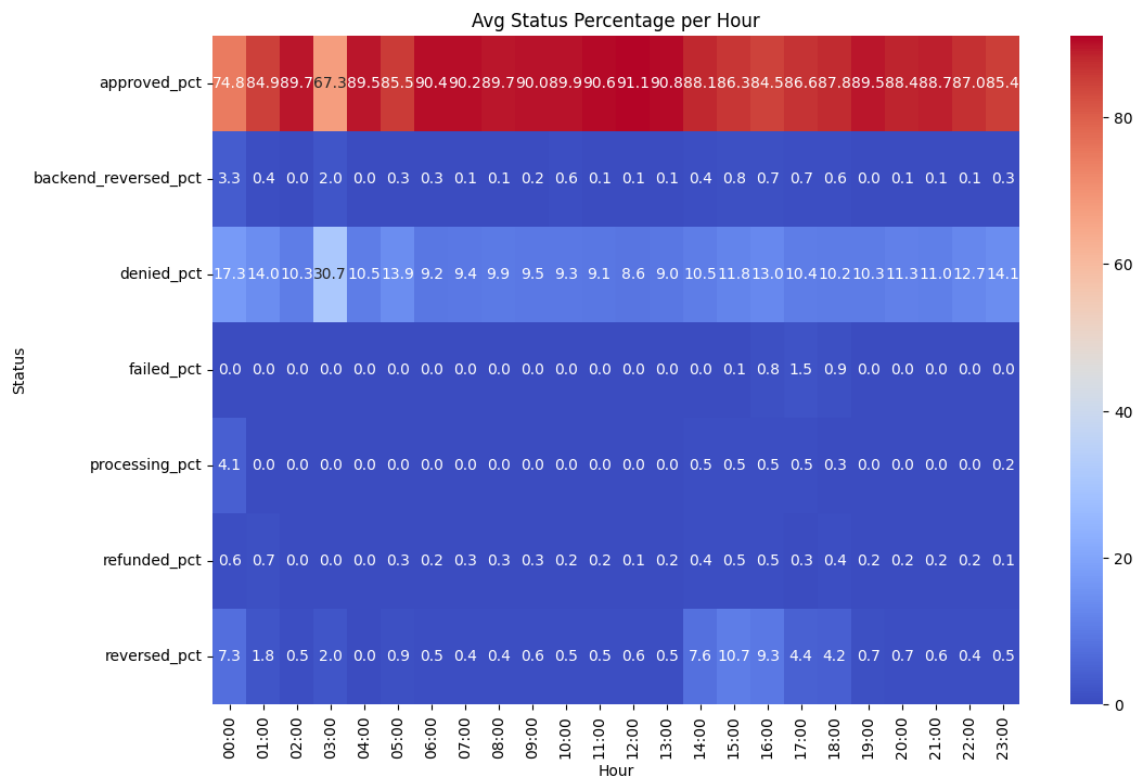


Figure 4: Alternative Status Percentage heatmap chart

From the analysis, we observed the following patterns:

- **Approved Transactions:** The majority of transactions are approved, peaking during the day from 10:00 to 16:00. This pattern suggests high business activity during these hours.
- **Denied Transactions:** There are consistent occurrences of denied transactions throughout the day, with a noticeable increase during peak business hours.
- **Failed and Reversed Transactions:** These occur less frequently but have spikes during certain periods, indicating potential issues or anomalies in transaction processing.

We have also identified the following potential anomalies:

- **Unusual Spikes**
 - Spike in denied transactions percentage, around 03:00,
 - Spike in failed transactions percentage, around 15:00,

Which could indicate risk management interventions or other issues.

Conclusion

The analysis indicates that the majority of the transactions occur during the day. With increased transaction volume on business hours, peaking from 10:00am to 4:00pm. By visualizing the transaction status percentage over time we are able to see potential anomalies that occur around 3:00am and 3:00pm, with increased percentages of denied and failed transactions. Further investigation is recommended to understand the cause of these spikes.