Monitoring Test – Gustavo Moraes

Objective

Challenge proposed by Cloudwalk during Monitoring Intelligence Analyst selection process. The objective of this test is to develop two stages:

- The first stage involves data analysis to identify potential anomalies in the dataset.
- The second stage consists of creating and implementing a real-time monitoring system (integrated with Microsoft Teams) to detect possible anomalies in the dataset.

Checkout Analysis (item "3.1 - Get your hands dirty")

To identify and highlight potential data anomalies, several statistical techniques were employed, including visual inspection, data aggregation (mean), the Interquartile Range (IQR using Q1 and Q3), and the Z-Score method. Throughout this analysis, AI tools were used playing a supporting role, assisting in code generation and idea development. For example, while the initial plan was to apply only the IQR method, AI also suggested incorporating Z-Score analysis. It's important to note that all AI-generated content and documentation were carefully reviewed and validated by me.

Raw Data Analysis:

- In Checkout 1, notable anomalies were observed at:
  - 8h, 9h, 15h, and 16h when comparing "today" and "yesterday"
  - 10h across "today", "yesterday", and "same_day_last_week"
  - 12h and 17h with "today" data
- In Checkout 2, "yesterday" presented frequent anomalies, and anomalies at 15h, 16h, and 17h were evident for both "today" and "same_day_last_week", particularly when compared to "avg_last_week" and "avg_last_month".

Average Data Analysis:

- In Checkout 1, anomalies begin to appear after 10h, continuing until around 23h. A similar pattern is seen in Checkout 2, though the anomalies appear more pronounced when compared with “avg_last_week” and “avg_last_month”.

Boxplot Analysis:

- In Checkout 1, “today” and “yesterday” show higher variability compared to “avg_last_week” and “avg_last_month”, indicating more frequent anomalies in recent days. The same is true in Checkout 2, with “yesterday” showing particularly wide variability. Notably, no statistical outliers were detected in either case.

Z-Score Analysis:

- In Checkout 1, anomalies were consistently present from 10h to 23h, with elevated Z-Score values. Checkout 2 displayed a more irregular trend, but anomalies were still detected at 15h, 16h, and 17h. Again, no values fell outside the typical outlier threshold.

SQL With IQR and Z-Score Analysis:

- Individual evaluation via SQL confirmed that Checkout 1 maintained more stable behavior, while Checkout 2 exhibited relatively more frequent anomalies. However, no extreme outliers were identified in either case.

The overall analysis indicates that Checkout 1 experienced higher-than-average sales from 10h to 23h, with consistent performance across “today”, “yesterday”, and “same_day_last_week”. In Checkout 2, a similar pattern is seen, except for a noticeable drop in sales between 15h and 17h. Despite these fluctuations, no severe outliers were found. In summary, sales appear generally higher than usual, with only minor anomalies requiring attention. Aside from these cases, the data remains within expected parameters.

Transactions Problem (item “3.2 - Solve the problem”)

To address the problem, a comprehensive solution was implemented using a Python notebook on Google Colab, the Isolation Forest machine learning model, and an API integration with Microsoft Teams via Flask and Ngrok.

The process began with data extraction and preprocessing to ensure compatibility with both the machine learning model and predefined rule-based logic. Afterward, the Isolation Forest model was trained, and the fixed rules were applied, enabling the system to detect anomalies and generate alerts along with corresponding visualizations.

Next, an automated Teams messaging system was developed using custom-built APIs with defined URLs and endpoints. Each alert message sent to Teams includes detailed information such as:

- Time of failure
- Volume of failed, denied, and reversed transactions
- Indication of whether the alert was triggered by the machine learning model or the rule-based system
- Link to a graphic generated from the analysis

It’s important to point out that AI and Internet Sources were used throughout the development process. The use was purely academic and for assistance, with all information previously reviewed and (most of the time) modified by me. At the end of the project, a real-time analysis is performed using fixed rules and ML. In addition to delivering this information, the project shows a graphic and automatic messages on the Teams.