Hotel Booking Dataset from Kaggle

Project No2:

Static but Fully Updated KPIs Dashboard in Looker Studio

Project Management

This is a complementary project to 'Project No1: Leveraging Machine Learning for Cancellation Prediction: Report to Stakeholders'.

For detailed project management and automation information, please refer to the ***Full_Project_Explained.pdf***. This file is common to both projects and solves all the questions.

The reason for creating this extra file is to keep both projects' documentation short.

Tasks

Task 1: Data Preprocessing

- Script file: ***preprocessing.py*** (extracts ***hotel_booking*** raw data from PostgreSQL, transforms it and finally loads cleaned ***logreg_rf_data*** and ***dashboard_data*** to PostgreSQL).

Task 2: Generation of KPIs

- Script file: ***dashboard_dataframe.py*** (extracts ***dashboard_data*** from PostgreSQL, transforms it and finally generates cleaned ***hotel_market_segments.csv*** and ***hotel_kpis.csv*** which are stored on my local machine).

Task 3: Generation of Looker Studio (LS) KPI Dashboard

- Data files needed: ***hotel_market_segments.csv*** (for vertical bar charts) and ***hotel_kpis.csv*** (for scorecards).

- There are buttons with document links inside LS. For documents and button information, see the next section.

Documents

KPIs_Significance.pdf → It explains why the used KPIs are significant (LS button named 'Explain KPIs').

Report_to_Stakeholders.pdf → This is the report of Project No1. I place it inside KPI dashboard so that stakeholders can access the ML output report apart from the KPIs (LS button 'Machine Learning Report').

Full_Project_Explained.pdf → See section's introduction above (LS button 'Project Workflow & Automations').

Doc_KPIs_Looker.pdf → This is the project's general documentation file

Goals

- To deeply engage with hospitality KPIs.
- To create calculated fields outside of Looker Studio using Python, avoiding low-level dashboard coding and minimizing loading and updating times.
- To build a visually appealing Looker Studio dashboard.
- To semi-automate the process so that the user only needs to update the datasets inside Looker Studio.
- To present a real-life case study by building an end-to-end ETL data pipeline, using exploratory machine learning (feature importance) to identify and define hospitality KPIs.

Assumptions

- The company stores data in a local PostgreSQL database and I have permission to extract and load the data.

- I can remotely access and boot my work PC when needed.
- The company provides Looker Studio KPI dashboards to its customers at agreed time intervals, such as monthly.

Preprocessing

The most important preprocessing steps are described in the Preprocessing section of 'Project No1: Leveraging Machine Learning for Cancellation Prediction: Report to Stakeholders'. The ***preprocessing.py*** file is common for both projects because it generates ***logreg_rf_data*** and ***dashboard_data***. Of course, ***dashboard_data*** is the dataset of interest for this project, and for this reason, I include comments in ***preprocessing.py*** related to this dataset. However, there are no important preprocessing things to mention about this dataset.

KPIs Creation

I calculate KPIs based on both exploratory machine learning (Project No1) and hospitality domain knowledge. For example, I include financial KPIs in addition to KPIs related to booking cancellations, to provide a complete picture of strategic KPIs to the stakeholders. The ***dashboard_dataframe.py*** file includes the scripts for KPI calculation. I store the KPIs in two tabular structures (***hotel_market_segments.csv*** and ***hotel_kpis.csv***) locally on my machine without loading them into the local PostgreSQL database.

Looker Studio Dashboard

The Looker Studio dashboard displays the pre-calculated fields included in ***hotel_market_segments.csv*** and ***hotel_kpis.csv***. The dashboard's KPIs have been created based on stakeholders' preferences and strategic planning.