Cross Solution Data Analysis: Towards Consistent KPIs Across Solutions

Capstone Project Summary

By Tunay Tokmak

In partial fulfilment of the requirement for the degree of Master of Business Analytics

Submitted to
Central European University
Department of Economics and Business

Supervisor: Zoltan Toth

Vienna, Austria

June 2023

Table of Contents

Summary

The client organization offers standalone document management software solutions but lacks a holistic evaluation of their performance due to disintegration across products. The project aimed to provide feedback on current data structure, propose a brand-new robust data structure and develop interactive Tableau dashboards for metrics and KPIs visualization as well as the generation of the metrics and KPIs which integrate the standalone products. As result, a brand-new data structure has been devised enabling the generation of holistic metrics and KPIs across solutions, and Tableau dashboards that showcase the generated metrics and insights are presented to the organization.

Problem Definition and Objectives

The client organization is a provider of document management software solutions, catering to a wide array of clients. Its portfolio consists of several products, such as document capturing, text extraction, workflow management, and document scanning. However, the existing approach to data collection from these standalone products makes it impossible to evaluate their performances in an integrated manner. In this context, the objectives are as follows:

1.Provide the company with valuable feedback on data quality and come up with a brand-new robust data structure which enables the generation of integrated metrics and Key Performance Indicators (KPIs) across solutions.

2. Design and develop interactive and reproducible dashboards that showcase the generated metrics and KPIs across multiple software solutions using Tableau visualization tool.

3.Create performance metrics and key performance indicators such as lead times, delay times, throughputs, change in cycle times etc. to increase value and capture risks in the business operations.

Data

The client organization's datasets come from two main software solutions which are invoice capturing and document management. Unfortunately, the collected data suffers from poor quality, necessitating the creation and utilization of an artificial dataset during the Tableau visualization process. To generate this artificial data, various Python libraries, including NumPy, Pandas, Random, Datetime, and Faker, were employed. The artificial dataset partially represents the underlying data, incorporating processes and steps within business operations in original dataset. However, it is important to emphasize that the time-related variables within the artificial data do not reflect reality due to the impossibility of identifying lead times and delay times from the original dataset.

Outcomes

Feedback

As a consultant, valuable feedback regarding the existent data structure and quality is given to the client through exploratory analysis of the data in reproducible Jupyter notebooks. The analysis highlighted necessary improvements in the data structure, which will enable clients to generate KPIs and make data-driven decisions across products.

Data Structure

A brand-new robust and holistic data structure is devised, and an artificial dataset has been generated to demonstrate the suggested data structure’s feasibility. This new data structure allows for integration of multiple processes in a sequential manner to reflect the business operations in a dataset. The artificially generated dataset is accompanied by a data dictionary and reproducible source code to ensure the reproducibility and feasibility of the project. By adopting this new data structure, the client organization can gain insights about business operations instead of individual processes and this might result in cost reduction by applying different strategies like resource allocation, employee training, infrastructure change etc.

Metrics and KPIs

With the utilization of the new data structure, efforts are made to generate meaningful metrics and KPIs across solutions. These metrics and KPIs focused on an integrated view of the document management solutions as well as standalone analysis.

Tableau Workbook

As a visual representation of the generated metrics and KPIs, an interactive and interconnected Tableau workbook is created. This workbook consists of 3 connected dashboards which correlate processes with underlying business operations, and it displays generated performance metrics.

Limitations

This project has a couple of limitations that should be considered. Firstly, the data sources provided by the organization are limited to invoice capturing and document management solutions, solely two solutions. As a result, the generated dashboards may only offer insights specific to these solutions and may not provide a comprehensive understanding of the organization's overall performance or other software solutions. It is crucial to note that the metrics and KPIs derived from these dashboards might not be directly applicable to other software solutions within the company. Secondly, due to time constraints, user testing and responsiveness testing across different devices, such as mobile devices, are not have been completely conducted. Therefore, it is recommended to allocate time and resources for testing and optimization before any production phase.

Knowledge Gained

- Data quality is the most important factor in the success of any data-driven project. To ensure the quality of data, an analyst must consider integration with other data sources, format and clarity of the datasets. In the process of artificial data generation, I have upgraded my reporting skills as an analyst and the skill of ensuring data integrity.
- Tableau is a very powerful software, and this project provided me with the skillset to design and develop interactive and visually captivating dashboards.
- The project improved my capabilities in communicating with stakeholders in an effective and efficient manner thanks to progress reports, weekly/bi-weekly meetings and reporting.
- I have deployed brand-new Python libraries as well as coding techniques which contributed to my Python knowledge immensely.
- I have gained domain knowledge in document management systems as well as generating and visualizing key performance metrics and indicators.