Business Intelligence and Operational Analytics at no Cost for a Small Business A Case Study Capstone Project Public Summary Lilla Tagai - CEU Business Analytics 2018

My capstone project's goal is to create a reporting dashboard for customers and internal users for a small IT focused company and help discover with technology will serve the business' interests the best.

I worked with the startup company Talk-a-Bot who run a chatbot agency.

The company has extensive amount of production data stored in PostgreSQL databases. Due to rapidly increasing demand for business intelligence, they need an automated and cost-effective reporting solution of the most important KPIs. This will be used by both internal teams and offered as a service to clients.

The capstone project is a case study of options and concerns investigated while looking for the best fitting business intelligence tool.

During the course of the project several solutions were tried and discarded due to a number of reasons (this is elaborate in the Project Technical Discussion). Priorities also shifted when the business realized the effort required for a proper, all-customer reporting.

The project completed the discovery phase for a future analytics solution. During the work I had the opportunity to gain valuable hands on skills and experience.

Goals

The primary goal of the project was to provide reporting via simple standard reports that don't require elaborate development.

The secondary goal evolved during the project, it was to make a decision about the long-term analytics solution.

Benefits

The KPI dashboards serve double purpose: customers are offered it as a service, and they are used by the sales team to promote the product. It's important to define KPIs that best show the value of the product.

The bot's goal is in most cases to keep visitors interested until the process leads them to a human operator, so instead of reducing the workload of human operators, the bots are opening up a new sales channel.

Data infrastructure constraints

Infrastructure is a crucial aspect of any data solution. In our case, the central databases hold operational data, so that database performance can impact the user experience. Reporting or pulling data in any form should avoid having an impact on the performance of the tool.

The industry standard solution for this issue is instead of querying production data directly, is to collected data in a dedicated analytics database. Talk-a-bot decided to leave this effort for when they are ready to move their data to a cloud-based solution. We've decided for creating SQL queries that run once per day during the night. This minimizes the performance impact.

About the data

Each customer has their separate relational database, with table structure that evolved over time. I was provided to a subset of the databases in order to avoid any impact on the production databases. This allowed me to focus on understanding the data structure and defining the most optimal KPIs.

Project phases

Technology selection Definition of KPIs Dashboard design and development SQL queries

Technology selection

Solutions presented and considered

- 1) Tableau not selected due to licensing reasons
- 2) MS Power BI not selected due to infrastructure reasons. Used for prototyping
- 3) SQL queries and web interface (bulma.io framework) short term solution
- 4) Google Analytics Google Cloud selected as long-term solution

I am including the summary of tools used in the Project Technical Summary.

The backend will remain the same set of relational (PostgreSQL) databases and will be queried using SQL queries directly. For the KPI measurements I created these queries. For trendlines I created time series data by collecting these into a new stat table.

Visualization tool

The original prototype of the dashboard was developed using Microsoft Power BI. In terms of front-end capabilities, the tool was a good fit. The original plan was to use one common reporting template, and switch data sources between each client's data.

The team decided at the end for a self-developed web frontend to avoid vendor lock-in. The created Power BI prototype was used as a basis for the frontend of the web interface.

Defining of KPIs

I defined the KPIs based on the feedback of Talk-a-bot's management team.

The tool's goal is to keep the users chatting until they can be transferred to a human operator. This requires the bot to keep getting better at giving meaningful answers, so that the conversation can progress.

A few examples of KPIs:

- Number of visitors
- Number of subscribers and unsubscribes
- Number of interactions
- Number of messages (in/out)
- Number of helpdesk tickets

Dashboard design and development

The dashboard's initial design was created using MS Power BI. I focused on providing clear, concise measurements that can be interactively used by the end-users, and I tried to predict the questions the users will have. I learned how to model and visualize data with MS Power at an advanced level.

SQL queries

The presented data is pulled from the production databases via embedded SQL queries. I created three versions of each KPI, a same-day, a 30-days one and a grand total. For time series I created a new table, so that trend lines can be visualized.

Learnings

My most important learnings during my capstone project were

1. Operational requirements and needs change

As I mentioned in the first paragraph, the needs and priorities of the business stakeholders changed over time. This is a natural evolutions data analyst have to calculate with.

2. Infrastructure

Finding and setting up the right infrastructure is a crucial point in the lifecycle of an analytics project. Having a database for analytics purposes in necessary to enable commercial BI tools and avoid impact on production databases.

3. Data understanding and documentation

Creating data dictionary and entity relationship diagrams is essential. To be able to start analyzing the data I had to understand structure and content of fields. No documentation existed, I obtained verbal information and created entity relationship diagrams.

4. Data modelling

Data modelling is the majority of work when designing reporting. Without thorough data modelling, the data won't provide full benefits to the users.

5. SQL is fundamental

SQL is a fundamental BI tool one will have to reach back to when working with data. Data is still very often stored in relational databases, and SQL skills are an excellent basis to understand the query languages of MS Power BI or Tableau. The industry considers SQL skills as given baseline for data analysts.

6. Dashboard design

Less is more. It takes practice and good knowledge of the data and the business goals to create dashboards that provide users with just the right amount and type of information.