Business Intelligence

Project Assignment Handout

Spring Semester 2020/2021

This handout details the rules for the mandatory practical project for Business Intelligence.

Project Intermediate Delivery Deadline: April 3rd

Project Final Delivery Deadline: June 4th

Project Group: 3 to 5 students.

Project Summary

The project for this class is meant to reinforce the conceptual knowledge that you'll be acquiring throughout the course, asking you to apply core concepts of business intelligence, tools and methodologies to deliver an end-to-end Self Service BI solution to support strategic, tactical and operational business decisions, for whatever information problem that you choose to address.

During the course of the practical project, each group is expected to develop a proper Dimensional Model (using **Power BI Desktop**), a Report and Dashboard (using **Power BI Desktop** and **Power BI Service**) to support the different level of business decision makers. Additionally, you will have to write a .pdf report to describe, in detailed fashion, the work your group carried out.

Group Rules

The project should be done in a group of between three (3) to five (5) students; we consider this the ideal size for group work, and do not allow other sizes.

Should you constitute groups of size other than 3-5 members, you will be subjected to the following penalties:

> ➢ For each member in excess or below the stated size (3 to 5 members): penalty of **1.5 points** (out of 20 total points) in final project grade.

For example: should you choose to constitute a group of 7 members, your final project grade will be penalized in 3.0 points (out of a possible maximum grade of 20 points).

The only allowed exceptions that will be made regarding group size will be for situations beyond the control of the group members (for example, one of the students dropping out of the course) and will be evaluated on a case-by-case basis.

You are allowed to have group members from different BI I classes, and the members can, of course, be participating in an entirely virtual (remote) fashion.

Source Data

Each group must choose a business/organizational problem that can be used as the foundation of this practical project. In most real-life cases, BI projects usually involve large sets of source data, from where you will be extracting the relevant data for your project. You can use different types of data sources, such as relational databases, excel files, flat files, online datasets, etc.. We expect that your source data should contain, at the very minimum, the following features (some of these will become clear as the course progresses, and you become familiar with business intelligence terms):

> ➢ Some sort of transactional records that can represent quantifiable facts, to be used in your future dimensional model; these need not be extensive in volume but should represent a few thousand records of business transactions. In terms of BI, transactional means data that is used to count, sum or otherwise keep track of quantities of something that the organization deeply cares about. A transactional view of football would mean match results (1 = win, 0 = draw, etc.), match points, number of goals scored, etc. Information about the history of football clubs is NOT transactional; however, financial details about player sales between clubs is transactional. If possible, select transactional data that has regular, "well-behaved" quantitative measures. In other words, select data that keeps track of quantities that are easy to understand, can be easily operated on (summed, averaged, etc.) and that display a reasonably regular pattern across time

> ➢ Enough attributes (characteristics) so that you can extract at least five dimensions from the source data, according to the dimensions criteria for the project. Select data that has viable options in terms of dimensional analysis: in other words, data that has the potential to be analyzed (filtered, grouped and broken down) by several dimensions. **The characteristics that you will be setting up as dimensions of your model should provide you with the ability to set up different hierarchies in the different dimensions.**

> ➢ If you provide a relational database, it should be in Microsoft SQL Server 2019 (or compatible) version, or able to be converted to this format by your group (it is up to the group to ensure this conversion).

> ➢ The quantifiable data that will make up the facts of your multidimensional solution should represent **several years of transactional history** (at the very least, more than one year).

> ➢ The submitted solution, to be run by your teacher, must also have clear and direct access to your source data, without requiring some sort of custom solution.

To help you in these tasks, we have below some public websites from which you may be able to gather open data for your project.

- http://dadosabertos.pt/ (Portuguese "index" site, that links to several open data initiatives)
- https://www.sns.gov.pt/transparencia/ (Portuguese health system open data)
- https://dadosabertos.turismodeportugal.pt/ (Portuguese Tourism open data)
- https://opendata.bnportugal.gov.pt/ (Portuguese national library open data)
- https://datasetsearch.research.google.com/ (Google's own dataset index)
- https://www.kaggle.com/datasets/ (a very popular and useful open data repository)
- https://github.com/awesomedata/awesome-public-datasets/ (organized by industry)
- http://www.who.int/gho/en/ (global Health Observatory data from WHO)
- https://registry.opendata.aws/ (open data from a cloud giant, AWS)
- https://catalog.data.gov/dataset/ (USA's government open data portal)
- https://www.kdnuggets.com/datasets/index.html (datasets for Data Mining, Data Science and Machine Learning)
- http://getthedata.org/ (you can ask data-related question, or browse previous answers)

IMPORTANT NOTE: Please do NOT USE as your source data a database, dataset or some other type of data that is usually employed in training / teaching / demonstration settings, as these typically have BI solutions already available. For example, do not use AdventureWorks, Northwind, WorldWideImporters, Chinook, Sakila, etc.

Project Deliverables and Solution Requirements:

Intermediate Delivery – 5 points (out of 20):

Your group must deliver the chosen **source datasets** and a **.pdf report** containing the following points:

i. Introduction.

ii. Presentation of Business/Organization (making use of statistical data and descriptive aspects).

iii. Identification of Business Needs and Problem (including the context behind the informational problem that the DW/BI system is going to solve).

iv. Description of the source data and discovery process. **Consider at least 2 external sources of information.**

v. Definition of, at least, **2 perspectives of analysis**. (ex: operational, market, competitive analysis).

vi. A **draft design** of the **dimensional model** (star schema is recommended) and in-depth description of the model. The model must feature:
    a. At least **5 dimensions**, one of which must be a Time / Date dimension.
    b. At least **3 hierarchies**, with **average depth of three levels**.
    c. At least **one Fact table**, featuring at least **1 measure for each perspective**.

The PDF report should have a **maximum of 15 pages** describing the previous points. Exceeding this number **will incur a 0.5 point penalty for each extra page**.

Final Delivery – 15 points (out of 20):

For the final version of the project, each group must deliver:

i. A **Power BI Report (.pbix file)** containing:
    a. All data sources integrated in **a single Dimensional Model** (following the requirements of the intermediate delivery).
    b. All defined **hierarchies correctly configured**, as well as **modelling requirements** (column data types, sort, etc.).
    c. Definition of a **Date dimension**.
    d. At least **one Calculated Column** using DAX.
    e. At least **two DAX Measures** for each perspective of analysis.
    f. At least **two DAX Time Intelligence Measures** (YTD, YoY, Same Period Last Year, etc.).
    g. At least **one KPI visual** for each perspective.
    h. At least **one Advanced visual/scenario** using machine learning and /or predictive analytics.
    i. At least one report page for each perspective of analysis.
    j. Implementation of some sort of **storytelling technique** (tooltips, bookmarks, etc.)

ii.   A **Power BI Dashboard** compiling the most important aspects of the report.

iii.   Intermediate Delivery **.pdf report,** additionally containing the description of the following points:

     a. **Data Integration, transformation and modelling**, explaining main transformation on data and Power BI tools used (Power Query, Model view, etc.).

     b. **Description of PBI Report** for each perspective - explain the calculated columns, measures, and most important technical aspects of PBI report pages.

     c. **Analysis and Discussion** of the PBI report and data (what does the data say?).

     d. **Critical Assessment** of the project.

     e. **Conclusion**.

The development, implementation and description of **extra work** is recommended and will account for **1 point** of you project final delivery (**0.25 points** each extra work development). We consider extra work any Power BI techniques not covered during class, such as M or DAX functions, new visuals from the market place, etc.

Consider key data visualization best practices when building reports and dashboards.

The PDF report should have a **maximum of 30 pages** describing the previous points. Exceeding this number **will incur a 0.5 point penalty for each extra page**.

Final Notice:

➢ **failure to deliver on time will incur a 0.5 point penalty for each late day** (for example, 4 late days will accrue a 2.0 point penalty).

➢ **Failure to comply with the delivery guide** (no proper naming of objects, duplicated or unclear files, improper folder configuration, etc.) will meet with a **once-only 0.5 point penalty**.

Questions and Clarifications

Should it be necessary, we will provide further clarifications and answers to questions from students, updating the respective Moodle project forum as appropriate.

Good luck with your project!