**Creation of Consulting Project module (Data Analytics)**

**Objectives for the Capstone module :**

* In response to a new target audience and adapted pedagogical objectives, design a "project" unit that summarizes their BI / Data analysis course.
* Based on a provided dataset (OLIST), propose a step-by-step methodology that will mobilize the skills learned in training and will put the learners in situation on their future job (business analyst, junior data analyst) to help a manager to make better decisions. For Example, "cost-benefit" analysis.
* Find intellectual but also playful challenges to engage the participants over the duration of the module. Be attentive to the learners' experience and look for a "practical" side to this methodology that they can use in the future.
* Note that this "guided" project phase will be the methodology for the last module of the program, this time on an "open" topic.

**Name: "Capstone Project Data Analytics / Data Consulting**

**Duration: 5 units (5 x 8-hour sessions, i.e., 2.5 weeks part-time)**

Wrap up and reflect on the whole Data Analytics journey, applying what we have learned throughout the course to one of the company's data set. Bring results as each group must create a presentation including a summary of the challenge, a description of the steps; a clear visualization or dashboard that displays insights on key questions, either numerically and graphically; a description of high-level conclusions and recommendations for the management. It will be a great way to practice all the tools, techniques and methodologies covered in the past and make everyone realize how autonomous they have become.

**Unit 1: Consulting Methodology and project setup**

*Lecture: We will explain during the lecture the consulting methodology and the process they should follow in order to treat any business case. Basically, we will try to decompose the methodology into the 5 units that will be studied during the Capstone business case. The 5 steps will correspond to the following:*

* Project presentation
* Understand and prepare the data (basic descriptive statistics)
* Create a Diagnosis using advanced descriptive statistics and statistical inference *(+ creation of case hypothesis*) and establish recommendations.
* Transform your recommendations into a presentation (or in this case beautiful dashboards) that will allow non-technical interlocutors to understand everything you did.
* Present the results in a fun story-telling way!

**(Exercise 1) Context and setup**

* Presentation of the consulting case and problematic:

*How to increase profit margin (focusing on customer satisfaction) while maintaining a healthy order volume?*

* Presentation of the Company (Olist) and their business model
* Presentation of the dataset
* Presentation of the 5 modules of the week so the students get a perspective on what they are going to do each day.

**(Exercise 2) Understanding Olist relational database.**

Assignment Overview

Perform an \*\*Exploratory Data Analysis\*\* (EDA) through the creation of the data schema of ‘olist` using (<https://kitt.lewagon.com/db>).

Assignment Overview

Import the data from Kaggle into Excel using Power Query. After reiterating the foundations of the past modules regarding Excel, we will clean the data trough Power Query.

This will allow us to:

* Check the different data types of the columns in our tables.
* Deal with missing and nan values.
* Use a basic statistical descriptive approach to understand the construction of the data.

Learning Objectives

* Import CSV files to Excel
* Use Power Query to import the data.
* Start using statistical descriptive analysis on different variables.

**(Exercise 3) Basic Exploratory Analysis (reviewing basic Excel tools)**

Assignment Overview:

Use the relationships between our different datasets to merge tables in order to have exploitable numerical tables that will be used to do descriptive statistical analysis on the following exercises.

* Relational database keys
* Table Merges
* Group by
* Basic Pivot Tables
* Pivot tables with merged tables

Learning Objectives

* Understand the links between our different tables.
* Create tables with the columns that will help us identify significant features for our analysis.

**Unit 2: Understanding and preparing the data with Excel (Data Analyze Ad-Din) and Power Query.**

*Lecture: We will present how to understand and clean any database using power query and the Excel Data Analyze Ad-in. These tools are perfect to treat databases since with a few clicks we can have powerful statistical analysis (correlations, descriptive analysis, covariance, histograms, sampling, etc). Also, Power Query is as well integrated into Power BI so this lecture will also prepare for day 4 so we can start with a more advanced analysis when we will create our interactive dashboards. Understanding the link between our different Olist tables will finally allow us to have a better business perspective and create pertinent tables adapted to our problematic.*

**(Exercise 1) Orders Analysis**

Assignment Overview

In this exercise, we will merge tables that will enable us to integrate the Data Analysis Add-in package in Excel. This way we can start making an exploratory analysis and use some statistical tools to see which are the most important variables in our data (ex: review score vs delay time).

* Merge tables
* Descriptive Analysis
* Correlations
* Distributions

Learning Objectives

* Get at ease with the Data Analysis Add in
* Be able to start looking for important variables in any database.
* Improve your basic Data Analyst reflexes by checking basic indicators such as the distribution, the correlation etc.

**(Exercise 2 - Optional) Reviews Translator using Excel Addon of google translate.**

Assignment Overview

We will Use the Add-in of google translate into Excel to translate the reviews and see if we can identify the main reasons of the low review score (it is always handy to know how to introduce external add ins into Excel).

**Unit 3: Statistical Inference for business analysis with Excel Data Analysis Add-in and the use of Power Pivot to treat relational databases.**

*Lecture: We will present the basics of linear regressions and how we can easily do them with Excel Data Analyze Ad-in. It is important that the students can make simple predictions for a business such as sales, profits, costs etc. Also, we will present how Power pivot will allow us to create the relational links between our tables in order to create Pivot Tables having 2 or more tables as a source (no need to merge the tables anymore).*

**(Exercise 1) Sellers Analysis**

Assignment Overview:

Using this new methodology of Power Pivot relational database, we will use pivot tables from multiples sources in order to make a deep analysis over the sellers. For now, we have seen that wait time is the most significant factor affecting the review score. Nevertheless, after translating our comments reviews, we saw that some sellers had bad reviews systematically.

Based on our analysis we might want to remove some sellers from Olist platform based on their persistent bad reviews so we can improve our profit margin

Learning Objectives:

* Basic Linear Regressions
* Creating a relational database in excel trough Power Pivot
* Using multisource pivot tables in order to NOT have to merge tables as we did in past exercises

**(Exercise 2) CEO request**

Assignment Overview:

In this exercise we will ask the students to make their last analysis with Excel. We will present some Revenue, cost, and review\_score\_cost so students will be able to integrate these variables in their descriptive/predictive analysis. By adding these features, they will find several recommendations (it will be up to them to think what the best solution could be in order to decrease costs).

Learning Objectives:

* Use everything that you have been tough until now and make your best recommendation to the Olist CEO.

**(Exercise 3 - Optional) Interactive dashboards with Power Pivot**

**Unit 4: Using Power BI to make statistical analysis and present your results through awesome dashboards.**

*Lecture: We will see how to do in a really summed up way everything we did in day 2 and 3 but now using only Power BI and then how to create amazing interactive dashboards. Today, companies are transitioning into Power Bi from Excel, nevertheless bot tools still very used and it is important to know how to treat data with both. They have both several things in common (such as Power Query engine), but Power BI is an optimized tool for dashboarding and interactive presentation of the data.*

**(Exercise 1) Loading and cleaning data Power BI.**

Assignment Overview

Review of the concepts already presented all along this bootcamp and start using power bi:

* Charge multiple sources – files – into Power BI (add power query instructions)
* Clean and organize your data (date adjustment, --)
* Manage the relationships between your data tables.

**(Exercise 2) Data exploration**

Assignment Overview

Now with the data in the engine, our objective will be to construct the same analysis in power BI. We are going to analyze our data, but first we should be able to create the metrics that would allow us to further explore and relation our data.

Learning Objective:

* Tables & Pivot tables (Matrix visual in Power BI) to understand our data.
* Feature engineering – creating pertinent metrics in Power BI for our business case.
* Dax vs Excel formulas
* Power BI time intelligence (adding a timetable to the database that connects with all tables and allows to filter matrix visuals with a common timestamp)
* Distance calculator (optional)

The hypothesis of this case study assumes that the score of the reviews is influenced by the delivery time of the orders (expected delivery vs actual delivery) distance, type of product, and many others.

We will explore in depth, with the help of the many visuals of power BI the relation between those variables:

* + Visualizing insights from our data (heath map exploration)
  + The time of delivery impacts our review score (scatter chart widget) – regression visualization- Can we visualize the same results as of our regressions?

**(Exercise 3) Dashboard’s construction**

Assignment Overview

Valuable insights are key to the development of business decisions and are an asset in today’s business environments. In this exercise we will apply all the necessary key performance indicators (KPIs) to create an interactive dashboard that will allow not only the final users to understand the business but also to help in the decision process making of any manager.

* + Create tooltips to complement the main visuals in PWBI.
  + Create linear projections based on the history performance of the business.
  + Custom made report based on the main hypothesis of the bootcamp.

**(Exercise 4 – Optional) Python integration in power BI ?**

**Unit 5: Communicate your recommendations as a data business consultant.**

*Lecture: We will emphasize on the methodology of a pitch/results presentation in order to make it clear, enjoyable and structured. For the moment there are two main possibilities that I am still considering; presenting directly on Power BI but making it look like a Power Point Presentation or making it in Jupyter and importing the interactive dashboards into the Jupyuter notebooks (Microsoft added that feature this year on power BI)*

**(Exercise 1) Story telling with Power BI**

Assignment Overview

Learning how to present your recommendations is almost as important as the recommendations itself. That is why here you will learn how to structure your findings using Power BI (or Jn) in order to have a concise, fun and clear presentation! So, continue working through your last findings (sellers, review score vs delay) and make a hot/ awesome presentation that will last 5 min to convince the Olist CEO of your recommendations.

Learning Objective:

* Learn how to do a story-telling presentation to present your results
* Create a presentation on PWBI / Jn

(Optional)

Feel free to extend your analysis to help improve Olist’s profit margin:

* Should Olist remove only repeatedly underperforming sellers, after it has a honeymoon period of a few months?
* Should Olist remove the worst performing products / categories from its marketplace entirely?
* Should Olist restrict pairs (seller, customer) between certain states to avoid delays?
* Should Olist acquire new sellers, with some cost assumptions to be suggested?