|  |  |
| --- | --- |
|  | Analyzing a few datasets related to an Online retail Sales company and helping them with insights |
|  |  |

Version: 0.1

Contents

[1. Background 2](#_Toc19534597)

[2. Work needed 2](#_Toc19534598)

[3. Datasets 2](#_Toc19534599)

[4. Objectives 2](#_Toc19534600)

# Background

This case is about an online retail organization based out of South America.

This online retail store contains many orders and these orders are linked to the customers, sellers, products and customer reviews data. Its features allows viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers.

After a customer purchases the product from the website a seller gets notified to fulfill that order. Once the customer receives the product, or the estimated delivery date is due, the customer gets a satisfaction survey by email where the customer can give a note for the purchase experience and write down some comments.

Attention

An order might have multiple items.

Each item might be fulfilled by a distinct seller.

# Work needed

This document contains list of the tasks we would like you to complete in order to evaluate fully your technical abilities. Please read thoroughly all tasks and in case questions arise, do not hesitate to ask us.

You have 4 working days to complete all the tasks. However, if by some reason you cannot get all results in full, don’t worry. The purpose of this exercise is to understand how you would approach similar assignments, what would be your way of thinking and how you would advise Philips colleagues.

You are to work in 5 tasks related to same data sets. Each requires some data handling, modelling and visualizations. We would like to get following outputs from you:

• All the syntax-code / actual code from your software of preference. Please don’t forget to include notes and comments within

• Numerical outputs of the final models and statistical calculations (if applicable for the specific task)

• Data files with the required outputs in csv and/or .xslx formats

• Any visualizations of the data you consider useful to support your work (interactive charts would be considered as an advantage)

• Any other materials that you consider useful would be welcomed as well

At the end of the period you would be required to present your work in powerpoint in front of the team or via Skype.

# Datasets

The data repository contains eight different files related to the below categories:

1. Customer related excel files
2. Order items related variables
3. Payments related
4. Sellers related
5. Products related

# Objectives

This online retail company is looking for some valuable suggestion from your end.

1. Some customers didn't write a review. But why are they happy or dissatisfied?
2. With purchase date information, predict the future sales information. What is the expected Revenue for 2019? What are the factors that you might want to influence which may increase revenue in 2019?
3. Identify delivery performance and find ways to optimize delivery times.
4. Discover the product categories that are more prone to customer dissatisfaction.
5. There are 71 product categories. However, the ecommerce platform wants to optimize their number of product categories and want to reduce by 75%. How will they do that and which all will be the categories that they should concentrate?

