Case Study
Data Scientist Marketing Intelligence

February 2018

Scenario

- You were given the task to estimate how many tickets will we sell during the next weeks in a given country and through a specific channel.
- You have as **input** the data of company sales during the last year, split in three tables:

Orders_date	
column	**content**
id	Order identifier
date	Date of the order
channel_id	Channel through which the order was done

Orders_tickets	
column	**content**
id	Order identifier
n_tickets	Quantity of tickets in the order
type	Type of product

Orders_country	
column	**content**
id	Order identifier
country_1	First country
country_2	Second country

- You can download the files from:
  - https://drive.google.com/drive/folders/1xZne26eY84b4U4Pch-5-hGD6QqiIxkUo?usp=sharing

Notes:

- One order can contain several tickets
- Some orders are assigned to two countries. If there is no value in the “country_1“, use the “country_2“

Task

- The **goal** is to build a **script** that takes this information and estimates roughly how many tickets we will sell from the current date and in the next 10 days.

- Feel free to use the methods / tools / languages you want!

- The expected **output of the script** is a csv file containing the following columns:
  - date
  - country
  - channel
  - n_tickets

- Please prepare a **short presentation** of your process including the following points:
  - How did you do it?
  - Which statistics / criteria / assumptions did you use / make?
  - What difficulties did you find during the process?
  - Please share with us a screenshot of your code ☺