Analytics & Insights Case Study

Talabat Egypt

## Introduction

Thank you for your time and willingness to take on this challenge. We hope it’ll be a fruitful experience and that you will enjoy it - regardless of the outcome. Good luck :)

This is a technical challenge that aims to assess how you tackle the various analytics/insights requests, from understanding, to executing, and finally to presenting.

You are provided a simulation dataset for our orders throughout the past 4 months. We will be asking for some insights. You should use the data at hand to answer all questions.

## Data Source

The provided Excel file contains multiple sheets that should act as your input tables:  
 - fct\_order  
- dim\_vendor  
- dim\_date  
Also you will find a Data Dictionary sheet that documents and explains the business meaning of each field in those tables.

## Questions & Requested Insights

### Tableau

*++ Please submit all insights in one Tableau report, Python Notebook or Excel report that includes all answers and calculations.  
++Our recommendation is to use Tableau as it will boost your chances.*

++ *If any calculations are done outside of the submitted Tableau file using Python or SQL, make sure to include them in a separate file.*

1. Provide a chart with hourly, daily, weekly, and monthly average number of successful orders, add cuisines filter then add your insights and what do you think,  
   Try to provide all graphs in one chart with a selection to choose between daily, weekly, and monthly instead of having three charts.
2. Around which week/month did the orders start dropping? Why?
3. What is the Fail Rate in Egypt? *Fail Rate* is % of unsuccessful/failed orders out of total orders. Is it improving or aggravating over the past couple of months? Why?
4. We need to run an AB test, and for a technical reason we cannot run it simultaneously. We need to run each scenario for 5 days at least to collect enough data. State the pros (if any) and cons of the following AB test schedules:
   * A test: 17 Jan → 21 Jan, B test: 22 Jan → 26 Jan
   * A test: 17 Jan → 21 Jan, B test: 24 Jan → 28 Jan
   * A test: 17 Jan → 23 Jan, B test: 24 Jan → 30 Jan
   * A test: 17 Jan → 23 Jan, B test: 17 Feb → 23 Feb
5. We need to improve Fail Rate. For this, we need to identify the Fail Rate Top Offending chains that we need to work on **(Fail Rate Top Offenders:** are the chains we believe are worsening our Fail Rate i.e. making it higher). What are the top 10 chains that you believe are the top offenders? What are the main 1-3 issues that each chain is suffering from and how can we work on them? (*How to calculate to which extent a chain is offending our Fail Rate is left up to you and is an essential part for grading this question. Also, will you take all 4 months provided in the dataset or just a recent snapshot? And why?*).
6. We have a 10,000 EGP budget and we would like to use it to reactivate/win back customers that churned. How will you define churn? Shall we target all churned customers (and send a small amount to each)? Or shall we be selective (so few customers with high amount each)? And if we will be selective, mention 3-4 criteria/conditions we should use to filter for the top churned customers that deserve our investment. Please provide a final list of those customers we will send the voucher to.
7. Are the most loved chains ((the 6 chains with highest orders) operating in the most important cities (the 6 cities with highest orders)? Can we use a visualization that instantly pinpoints which important cities lack which loved chains?
8. As a delivery company, we have an important KPI called CPO (Cost per Order), which means how much money on average does it cost us to deliver a single order, which is mainly rider and trip cost. The November CPO report came in from the Finance team and showed that both cities: Alexandria and Minya had almost the same CPO - which the Head of Logistics felt does not make sense and is now arguing with the Finance team. Can you use the data at hand to resolve this dispute and support one argument over the other? Who do you think is right? *(One proof could be enough. But two proofs will be more decisive in resolving the issue between the 2 departments)*
9. A failed order is a bad experience for sure. Same is an order that arrives late. We want to send a big "we're sorry" voucher to 50 of our users who've had the worst experience recently. How can we choose them? Decide and apply this to come up with this list of 50 users that need this “we’re sorry” voucher the most.
10. Affordability is any kind of discount we try to apply to further encourage customers to place orders. There are 5 types of affordability: vouchers, free delivery, restaurant discount, item discount, and gem. Is there a correlation between affordability and the number of orders on a given day? Which kind of affordability affects orders the most?

### 

### SQL

*++ The answer for each question is a T-SQL script that should run correctly assuming there is a database that has the same tables provided in the Excel file with the same structure/schema. If you want to be sure it is working properly you can create such a database on your side but no need to submit it.*

*++ Please submit all scripts in one .sql file. Use comments to specify which script belongs to which answer. Also use comments to provide the final answer for each query.*

1. We need a table to show the Fail Rate per Talabat week, sorted by Talabat week.
2. How many customers churned in December (i.e. ordered in November and did not in December)? Segment those churned customers into 4 groups: “Frequent & Consistent”, “Frequent”, “Consistent”, “Neither”. *Frequent* customers are the ones that placed at least 12 orders in total from Sep to Nov. *Consistent* customers are the ones who did not miss a month without placing at least 1 order (apart from Dec, because this is when they churned).
3. We want to track if our customers pay more with each new order. For this, we need to know: In how many orders did the customer pay more than their previous order? Ignore each customer's first order since we cannot know if it was more than the previous one. Please also write down the resulting number in a comment below your query.
4. We need to calculate the MTD (Month-To-Date) active customers for every day. This means for every day of the month; we need to know how many customers placed successful orders throughout that month until this day. **This value for the last day of each month will basically reflect all customers that placed successful orders during this month**. Output is 2 columns: date, MTD customers. What month has the highest MTD active customers by day 16? How many?
5. We need to do the same MTD calculation but for % of retained customers. A retained customer is one who placed at least one successful order last month, and then showed up again with at least one successful order in the current month. Retention Rate % = customers who placed successful orders this month and the month before, divided by all the customers that placed successful orders last month. We need to calculate the MTD retention rate for every day of the provided dataset. Output is 2 columns: date, MTD Retention Rate%. What month has the highest MTD Retention Rate by day 25? How many?

### Python **(optional)**

*++ Please submit all answers in a Python Notebook (.ipynb). If Notebook is not possible, submit a Python script. This question is optional, only if you’re good at it.*

1. We need to forecast the daily orders for Jan and Feb 2022. Please train a model to perform this prediction, stating the expected accuracy of such a model.

## Final Notes

### How to tackle questions?

* All the answers are within the data - **please avoid wild guesses**.
* If a certain question is not clear, feel free to make an assumption and act upon it.
* All questions are mandatory, but if time is tight, just try to do as much as you can.
* In case you finished early, coming up with more insights would boost your chances.

### What to submit?

* Each section specifies what is the expected submission.
* Ideally, you should submit 3 files: 1 Tableau, 1 SQL, and 1 Python Notebook file.
* If you did calculations outside of Tableau in the Tableau section questions, please add your effort to the SQL/Python files you will submit for the SQL/Python sections.
* Please include everything in a ZIP/RAR file with your complete name separated by underscores (eg: firstname\_lastname.zip).

### Deadline

* You have 7 complete days since receiving the case study. If you receive it on a Monday midday for instance, you should submit it the Monday after - by no later than 11:59 PM.

### Grading

We will tend to evaluate the outcome from many different angles:

1. Understanding matters: Your ability to correctly understand the requested insight and translate it into the required action/effort.
2. Implementation matters: How well you approach the problem and do correct calculation/coding/scripting that results in reasonable insights.
3. **Presentation/Visual Storytelling matters BIG TIME:** Your ability to visualize your insights neatly and beautifully on Tableau, in such a way that attracts the business user as well as deliver the insight in a simple understandable way. This includes choice of chart, descriptive text, …etc everything! Same goes for how neat and presentable your submitted Tableau file is.