**Jellysmack - Data Science and Analytics Test**

The technical questions in this test are based on real-life scenarios and reflect a few completed projects. For your information, the tools used by the Jellysmack Data Team are as follows:

* **SQL**: DBeaver Community Edition: <https://dbeaver.io/download/>
* **Python**: Jupyter Notebook / Pycharm / Visual Code
* **Data Visualization**: PowerBI software or matplotlib/plotly in Python

Attached you will find:

* a Python file to begin your script
* a text file containing the login credentials to our test database
* a pdf file outlining how to use an online development environment

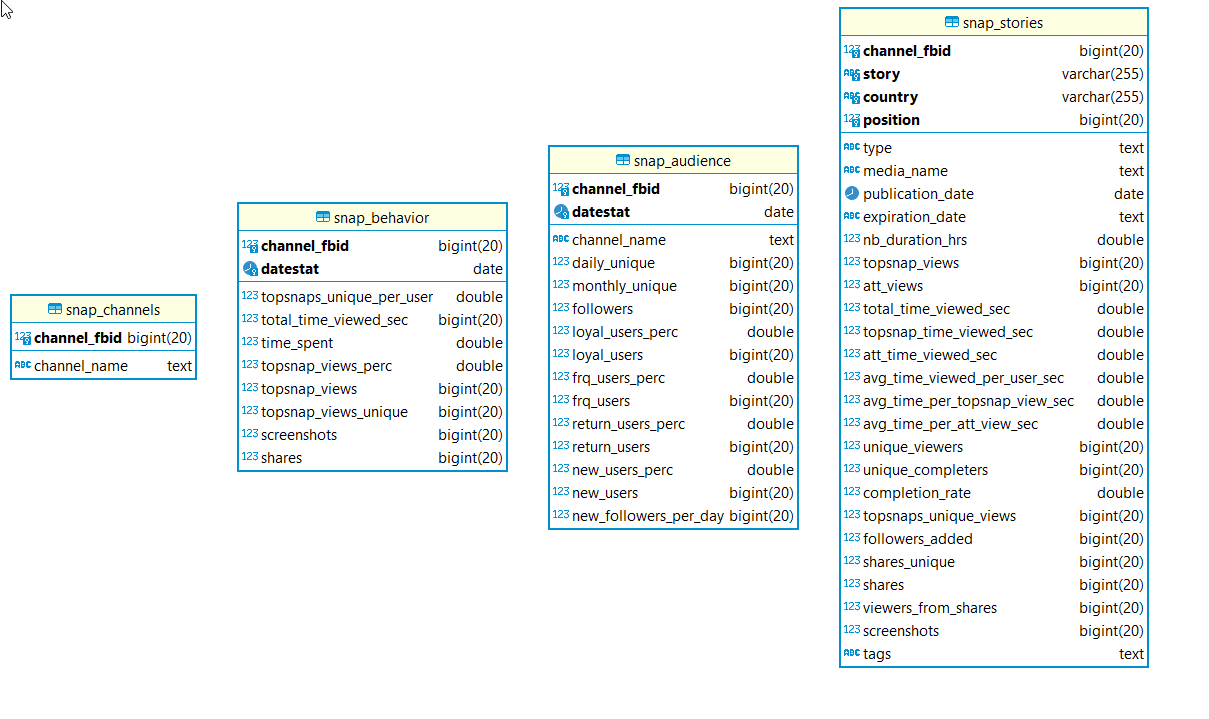
**Organizational Context:**

At Jellysmack, there is an Acquisition team whose job is to collect the raw data through various social network APIs. Those data are then stored in different formats (csv, json, parquet, SQL, etc.).

Our Data Analysis and Science Team examines this data on two levels:

* Data Analytics Engineers extract, clean, organize, transform, and consolidate this data to make it usable.
* The Data Analysts and Scientists will exploit this data, analyze it, and then communicate it to the Jellysmack teams via various media.

The Data Model you will be working on is as follows:



**Guidelines before getting started:**

*What is expected:*

* **Retrieve** data from our **MySQL** test database
* **Analyze** the data using **exploratory, descriptive**, and **statistical** techniques *(hypothesis testing, confidence interval, regression, ..)*
* **Visualize** the data
* **Provide feedback** on the findings of the analyzes and visualizations

*Evaluation criteria:*

* Analysis, visualization, and statistical reasoning consistency: interpretation and justification of results obtained
* SQL query quality and efficiency, as well as understanding of the Data Model
* Python libraries used effectively

**Jellysmack’s context:**

Followed on Facebook with about 300 million fans and 10 billion monthly views, Jellysmack is also present on Instagram, Twitter, and Snapchat.

This test will take you back to 2018, when **Jellysmack** was just getting started on **Snapchat**, through its first four channels and with the level of information we had at that time.

The analysis will center on the launch of **Accuse Moi Si Tu Peux**, the **RIDDLE**-themed channel. You will also be required to reflect on the theme: **BEAUTY**.

The **Snapchat** channel creation process:

Unlike other social media platforms, **Snapchat** requires channel creation to be approved to be released on their platform.

Creation of **RIDDLE**:

* On the 27th of June 2018, the Snapchat channel “**Riddle Me This”** (RMT) was approved and launched. This channel proposes various cool riddles and brain teasers in English, without subtitles, with an original design.
* To reach a French audience, the channel “**Accuse Moi Si Tu Peux”** was created. It got approved and launched on the **24th of August 2018**. This channel proposes the exact same content as the *Riddle Me This* channel but in French.﻿ When the *Accuse Moi Si Tu Peux* channel was launched, all *Riddle Me This,* French subscribers were automatically redirected to this channel.﻿

**Challenges**

1. The editor of **Riddle Me This** asked the Analysis team to determine the impact the launch of the *Accuse Moi Si Tu Peux* channel had on their **RIDDLE** theme.
2. The editor in chief of **BEAUTY** is wondering if they should launch a similar strategy for **Beauty Hacks** or **Beauty Wow** and in which language.

Please send your work to:

[data-tests-results@jellysmack.com](mailto:data-tests-results@jellysmack.com)

**Python MySQL Connection class :**

|  |
| --- |
| **import mysql.connector import pandas as pd from sqlalchemy import create\_engine  class JskDB:  """  Example of SQL connection class  """    def \_\_init\_\_(self):  # MySQL database credentials   self.host = 'jellybas.cxvglcwgczps.us-west-2.rds.amazonaws.com'  self.port = 3306  self.user = 'da\_ro'   self.password = 'UXDcu5ULheGEYJWF78Zy'  self.db = 'da2'   def create\_connection(self):  """Initialize MySQL connection"""  engine = create\_engine(f"mysql+mysqlconnector://{self.user}:{self.password}@{self.host}:{self.port}/{self.db}?charset=utf8",echo=False).connect()   try:  print(engine)  return engine  except Exception as e:  print(e)   # How to use  def my\_function():  """  open a MySQL connection, run your code and close the connection  """   # open connection  sql\_connection = JskDB().create\_connection()    # My code to do something in MySQL Database    # close connection  sql\_connection.close()** |

**Mysql Credential:**

host = jellybas.cxvglcwgczps.us-west-2.rds.amazonaws.com

port = 3306

user = da\_ro

psw = UXDcu5ULheGEYJWF78Zy

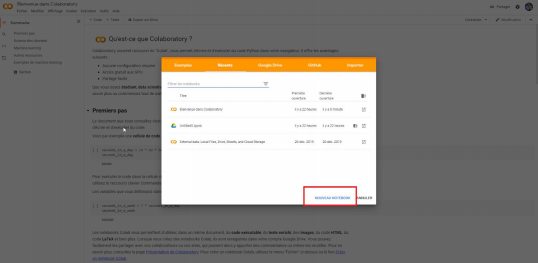
database = da2

**On-line development environment:**

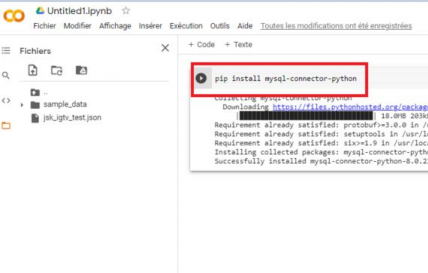
If you don't have a python development environment, you can use Google colab with the following prerequisites

1- **If you have a Google account,** click on the following link

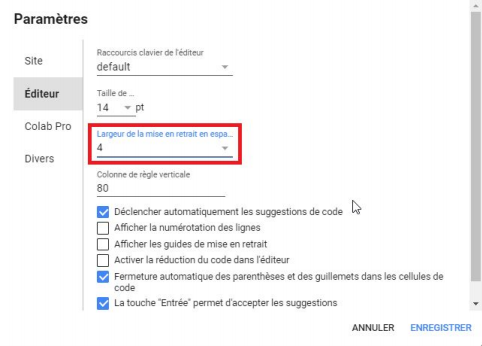
https://colab.research.google.com/notebooks/intro.ipynb#recent=true

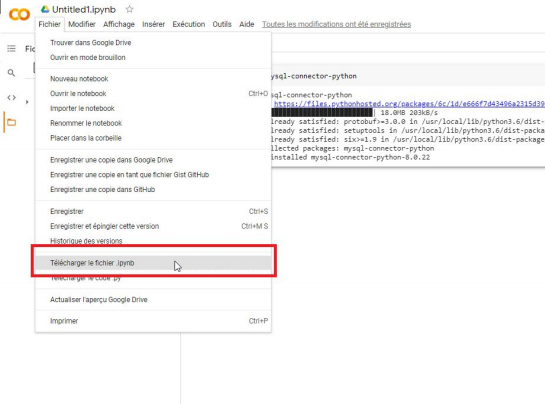
****

2- **Install the necessary libs**: You'll need the MySQL connector, which you can get by typing the following command into a cell: **pip install mysql-connector-python**

******

3- From the **Tools menu**, in the editor settings, change the width from 2 to 4

****

4- Once you have completed your work, **download** the ipynb file from the menu bar and send it to us ****

5- It should be noted that we also tested two visualization libraries:

matplotlib et matplotlib

**Glossary of Metrics**

Here is a glossary of metrics:

* **Story:** Name of the Story
* **Region:** Metrics region of the report. Should contain a "Global" entry in addition to local regions
* **Type:** Two values: Story or Snap. Story are Story totals, Snaps are individual snaps
* **Position:** Index value of the snap within the Story. Story index is null.
* **Media name:** Individual snap's name. Story media name is null.
* **Post time:** The time the post went live
* **Expiration time:** The time the post expired
* **Duration** (hrs): The length of time the post was live
* **Unique viewers**: Unique number of users that clicked into the Story or began loading a snap
* **Unique completers**: Number of unique users who completed all available content. For Stories, this is viewing through the last available snap. For Topsnaps, this is viewing through 99% of the runtime
* **Unique Topsnap per User**: The average number of Topsnaps in your story viewed by Snapchatters
* **Unique Topsnap views:** The sum total number of unique users for each Topsnap in the Story. For snaps, this is equivalent to unique viewers
* **Total time viewed** (sec): Total time spent in the Story or snap
* **Topsnap time viewed** (sec): Total time spent in the Topsnap(s)
* **Attachment time viewed** (sec): Total time spent in the attachment(s)
* **Average total time viewed per user** (sec): Total time viewed (sec)/Unique viewers
* **Average time per topsnap view** (sec):Topsnap time viewed/Topsnap Views
* **Average time per attachment view** (sec): Attachment time viewed/Attachment views
* **Topsnap views**: The total number of Topsnap views. Not unique, just total views.
* **Attachment views**: Total number of swipe up views. Not unique, just total views.
* **Unique attachment viewers**: Number of unique users that swiped up
* **Attachment conversion rate**: For Stories, this the average of all Topsnap attachment conversion rates. For Topsnap, this is Unique attachment viewers/Unique viewers
* **Subscribers**: Number of new subscribers added on this piece of content
* **Unique sharers**: Unique number of users who shared this piece of content
* **Shares**: Number of distinct edge shares (i.e. 1 person sharing to 3 friends = 3 shares)
* **Viewers from shares**: Unique number of users who viewed this piece of content via a share
* **Screenshots**: Number of screenshots
* **Males**: Number of unique male viewers
* **Females**: Number of unique female viewers
* **Unknown gender**: Number of unique unknown gender viewers
* **13-17**: Number of unique viewers of age 13-17
* **18-24**: Number of unique viewers of age 18-24
* **25-34**: Number of unique viewers of age 15-34
* **35+**: Number of unique viewers of age 35+
* **Unknown age:** Number of unique viewers of unknown

# **Audience & Behavior Analytics**

* **DAILY UNIQUES:** Number of Unique Snapchatters that engaged with your channel per 24 hours.
* **MONTHLY UNIQUES:** Number of Unique Snapchatters that engaged with your account per 30 days.
* **TOTAL UNIQUES VIEWERS:** Number of Unique Snapchatters that engaged with your account over any selected time period.
* **DAILY CORE USERS:** Number of Unique Snapchatters that engaged with your channel at least 3 out of the last 7 days.
* **TIME VIEWED:** Amount of time on average that Unique Snap Viewers spent viewing your account.
* **ATTACHMENT CONVERSION:** Percentage of Unique Snapchatters that swiped up on snaps with articles or videos attached
* **UNIQUE TOPSNAPS PER USER:** Average number of Topsnaps in your story viewed by Snapchatters.
* **UNIQUE TOPSNAP VIEWS:** Total number of Unique Topsnaps viewed by all Snapchatters that engaged with your content.
* **TOTAL TIME VIEWED:** Total number of minutes Snapchatters spent on your content (Topsnaps and attachments)
* **USER LOYALTY:** How frequently Snapchatters who view your channel return to your account in a given weekly period. Data is shown for the average of the last 7 days :  
  - New users : 1day/Wk  
  - Returning users : 2day/Wk  
  - Frequent users : 3-4day/Wk  
  - Loyal users : 5-7day/Wk

# **Story Analytics**

* **UNIQUE VIEWERS:** Total number of unique Snapchatters that engaged with this Story.
* **SUBSCRIBERS ADDED:** Total number of new subscribers to your account added when this Story was live.
* **TOTAL VIEWERS:** Total views of Topsnaps, articles and videos from this Story.
* **TOPSNAP PER USER:** The average number of Topsnaps in your Story viewed by Snapchatters.
* **TIME VIEWED:** Amount of time on average the Unique Snapchatters spent viewing this Story.
* **ATTACHMENT CONVERSION:** Percentage of Unique Snapchatters that swiped up on snaps with articles or videos attached.
* **SHARES:** Total number of shares of snaps from this Story.
* **SCREENSHOTS:** Total number of screenshots of snaps from this Story.

**Available data**

**Snap**

Also referred to as **Topsnap**

Looping video / video that auto-advances

**Attachment**

A piece of content that viewers swipe-up into from a Snap. Attachments can be articles, long-form videos, polls, votes, or quizzes. Also referred to as: long-form or bottom snap.

**Story**

The complete collection of Snaps and Attachments that you publish is called a Story. Also referred to as: edition (Publisher Story) or episode (Shows, Originals).

|  |  |  |  |
| --- | --- | --- | --- |
| snap\_behavior | snap\_audience | snap\_stories | snap\_channels |
| channel\_fbid  datestat  topsnaps\_unique\_per\_user  total\_time\_viewed\_sec  time\_spent  topsnap\_views\_perc  topsnap\_views  topsnap\_views\_unique  screenshots  shares | channel\_fbid  channel\_name  datestat  daily\_unique  monthly\_unique  followers  loyal\_users\_perc  loyal\_users  frq\_users\_perc  frq\_users  return\_users\_perc  return\_users  new\_users\_perc  new\_users  new\_followers\_per\_day | channel\_fbid  story  country  type  position  media\_name  publication\_date  expiration\_date  nb\_duration\_hrs  topsnap\_views  att\_views  total\_time\_viewed\_sec  topsnap\_time\_viewed\_sec  att\_time\_viewed\_sec  avg\_time\_viewed\_per\_user\_sec  avg\_time\_per\_topsnap\_view\_sec  avg\_time\_per\_att\_view\_sec  unique\_viewers  unique\_completers  completion\_rate  topsnaps\_unique\_views  followers\_added  shares\_unique  shares  viewers\_from\_shares  screenshots  tag | channel\_fbid  channel\_name |