**Test case №1:**

We extracted data from GCP. The dataset contains 4 million rows, but we were only able to extract them in small portions, about 40,000 rows in CSV format. In total, we ended up with 110 files. The files are named as follows:

data\_000000000001.csv

data\_000000000002.csv

…..

data\_000000000010.csv

data\_000000000011.csv

……

data\_000000000100.csv

data\_000000000101.csv

….

data\_000000000110.csv

Please suggest a method for uploading all these 110 files using Jupyter Notebook (or another) and merging them into a single dataset for further processing.

**Test case №2:**

You can find the file in this folder, which contains Linkedin profile scrapes of individuals that were part of an outreach effort. The goal is to get people to purchase a ticket to an event and ultimately join the organization as a member. Since the same method of profile scraping will be used in the future to identify the best leads, it's important that patterns that exist in the content of the profiles be identified.

The column 'Result' has one of six outcomes for each person:

-Purchased membership: The ideal outcome, bought an event ticket and a membership

-Purchased event: A positive outcome but not ideal

-Approved: They were approved to come to an event but did not purchase

-Sent personal f/u: Conversation was started but ended

-Out of sequence: Conversation was started but ended

-2021 Member: Someone who was a previous member of the organization

In order to better focus outreach efforts in the future, we need to determine which characteristics show a higher probability of purchasing a membership and/or purchasing an event ticket.

Please provide an analysis that identifies what traits, characteristics, patterns, etc. can be used to indicate a higher probability of a positive/ideal outcome and please describe your methodology.