

Applecart Commercial Analyst Exercise

About Applecart

Applecart's social graph technology allows us to identify and track relationships between individuals. By leveraging connections people have with each other, Applecart can help companies perform effective targeted messaging. For this exercise, you will be utilizing the social graph for a commercial application to analyze target universes as well as formulate a business plan. A target universe is a list of people we generate who are related to an initial group of people our client provides us. This list provides new leads the client can reach out to. For example, if a client gives us a list of people who bought one of their products recently, we could generate a target universe consisting of coworkers of people in that list.

Background

Salamander is a small but ambitious car insurance company based in the Midwest region that is looking to expand its business. As a pilot program, Salamander gave Applecart a list of recent customers in the Ohio area, from which we provided them two target universes as lead generation:

1. One target universe contains **first-degree connections** of the recent customers: people who, based on our social graph, are friends, coworkers, colocated (household members), or classmates of the recent customers.
2. The second target universe contains a randomly generated list of people.

Using these lists, Salamander launched an online ad campaign, sending online advertisements to people in both lists, and measured who clicked on the ads and who ended up buying one of their products. The results have come back and Applecart would like to see an analysis of the results from the first round as well as a proposal for improving future iterations of campaigns.

The Data

Attached are several datasets:

- **Recent_purchases.csv:** a list of people who have recently purchased one of Salamander's products. For example, in this row:

uid
42966404

There is a person with uid (Unique ID) 42966404 who recently purchased one of Salamander's products. A uid is a way of uniquely representing each person. In this exercise, the uids are represented the same across datasets, so a row in one dataset with a uid of 42966404 contains information on the same person as a row in a different dataset with a uid of 42966404.

- **People.csv:** a list of people Applecart has from Ohio containing demographic information including income, age, gender, race, religion, and ethnicity. For example, in this row:

uid	Family Income Detector	Gender	Birth Year	Race	Religion	Ethnicity
42669905	125000	M	1968	Eastern European	CATHOLIC	CZECH

There is person with uid 42669905 whose family has a yearly income of \$125,000, is male, was born in 1968, and is Eastern European, Catholic, and Czech.

- **Graph.csv:** a list of people's connections to each other and the type of connection. For example, in this row:

soure_uid	sink_uid	edges
1226148742	1834532311	neighbors

There is a person with uid 1226148742 who is a neighbor of the person with uid 1834532311. The graph is **bidirectional**, meaning that there is also a row where the source and sink uids are switched:

source_uid	sink_uid	edges
1834532311	1226148742	neighbors

- **One_degree_target.csv:** the target universe consisting of first-degree connections to people in the recent purchasers list, as well as the results of the ad campaign done on them. For example, in these rows:

uid	ad_action	buy_action
48525093	did_not_click	
42986174	clicked	bought_car

A person with uid 48525093 did not click on the online ad and did not buy a car insurance product, while a person with uid 42986174 clicked on the ad and bought a car insurance product. It is possible that a person did not click on an ad but still bought the car insurance product.

- **Random_targets.csv:** the target universe consisting of a randomly selected group of people, as well as the results of the ad campaign done on them. The format of the list is the same as one_degree_target.csv above.

The Project

Please send back to us the following deliverables:

- **first_round_analysis.pdf:** An analysis and comparison of the results from the one degree target universe and random target universe:
 - Which one yielded better results? What metrics did you use to evaluate which results were better?
 - What demographics are most strongly correlated with better results?
 - How do the results differ across different types of first-degree connections?
 - What issues did you identify with the data that Salamander sent you?
 - If you had more time to work on this project, what further analysis would you conduct?
- **business_strategy.pdf:** A business memo (around 3 pages) intended for internal distribution to Applecart. This memo should contain detailed and organized action items for Applecart to improve its follow up experiments, including:
 - Suggesting additional data Salamander should collect in its campaigns that would help them better measure success
 - Coming up with new Key Performance Indicators (KPIs) based on this additional data, explaining how each KPI is computed and why it is beneficial for Salamander
 - Coming up with an additional set of KPIs for Applecart to internally track how well its targeting strategies are working and how well the graph is performing
 - Suggesting how Applecart can improve the design and workflow of future campaigns so that it can more systematically track their findings and are capable of capturing the new KPIs you created. Consider marketing strategies, experimental design, platforms, and pipelines that would provide more streamlined and quantitative analysis
 - Any additional questions you would ask Salamander and/or Applecart that would help you make your recommendations
- **sql_commands:** A list of the SQL commands you used to conduct the first round analysis (make sure they are in an easily accessible format, like a notebook, pdf, or word document)

Suggestions

- It is encouraged that you do not work on this project for more than 6 hours. Spend around 1-2 hours understanding and ingesting the data, 2 hours analyzing the first round results, and 2 hours writing a memo suggesting improvements to the follow up experiments
- Consider all aspects of the consumer experience, from impressions to retention, in your business strategy memo. Be creative and detailed!
- Although the size of this dataset is small enough that it could be analyzed in Excel, the datasets that you will typically work with at Applecart are much larger and cannot be handled by Excel. As such, if you need to manipulate the data at all, you will need to do so using SQL (or at least let us know which SQL query you would use. If you want to use Excel to create graphics, that is fine.
- In order to complete the project, you will need to host the datasets that we send you. ***Applecart uses Databricks as its primary platform for SQL queries, so we highly recommend using Databricks for this exercise!*** Here are some links to get you acquainted with Databricks:
 - Overview of Databricks Community Edition (the free version of Databricks): <https://databricks.com/product/faq/community-edition>
 - A guide for getting started with Databricks (ignore Step 4, “Create a table,” and instead refer to the bullet point below on how to load data into Databricks): <https://docs.databricks.com/getting-started/quick-start.html#>
 - A guide for how to load data into Databricks:
<https://docs.databricks.com/user-guide/tables.html#create-table-ui>

You have 48 hours to complete this exercise. If you have any questions, please contact Alex Jeffers at alex@applecart.co