**Background**

Applecart’s social graph technology allows us to identify and track relationships between individuals. For this exercise you will use demographic data associated with the social graph in order to derive insights about which people are related to who.

Please do not spend more than 2-3 hours on this project.

**The Dataset**

Attached is a dataset of 34,037 rows with the following fields:

* AC\_Sink (a unique ID representing a person)
* AC\_Source (a unique ID representing a person)
* Sink\_gender (the gender of AC\_Sink)
* Sink\_YOB (the year of birth of AC\_Sink)
* Sink\_race (the race of AC\_Sink)
* Source\_gender (the gender of AC\_Source)
* Source\_YOB (the year of birth of AC\_Source)
* Source\_race (the race of AC\_Source)

Think of this as a crude version of the social graph. Assume that each row represents a connection between AC\_Sink and AC\_Source. So for example, if a row had AC\_Sink=123, AC\_Source=456, sink\_gender=M, sink\_YOB=1960, sink\_race=White, source\_gender=F, source\_YOB=1970, source\_race=Hispanic, this row is telling us that person 123, who is a white male born in 1960 is related to person 456 who is a Hispanic female born in 1970.

The other important thing to know about this dataset is that each relationship is represented twice. So for example, there is one row that is ac\_sink=123, ac\_source=456 AND another row where ac\_sink=456, ac\_source=123.

**Project**

Please analyze how demographics shape social relationships. At a minimum, please answer the following questions:

* Are people of the same gender more likely to be related? How much more likely or not likely?
* Are people of the same race more likely to be related? How much more likely or not likely?
* Are people of the same age more likely to be related? How much more likely or not likely?

Feel free to conduct additional analysis as well! Examples of additional analysis to conduct include:

* Are certain races more likely to have relationships within their race?
* How would the “age” analysis change if you used a range rather than an absolute number?
* What demographics are over or under represented in this dataset?

IMPORTANT: Although the size of this dataset is small enough that it could be analyzed in Excel, the datasets that you will typically work with 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.

**Output**

A written memo that is between half a page and one page with the findings from your analysis. The memo may also contain an appendix with as many charts or visualizations (or links to those charts/visualizations) as you wish to include. The memo may also contain a section for next steps you would have taken if you had additional time.

Additionally, please include a separate file with any SQL code that you used or would have used to manipulate data for this analysis.

Send your output to [ezra@applecart.co](mailto:ezra@applecart.co)