

STAT 426 Project Proposal

1. “How much money will I spend on my Apple Credit Card across different categories in January – March 2021?”

- a. Different categories are defined within the data in a “Category” column. (The different Categories before I created new ones were ['Other', 'Gas', 'Restaurants', 'Debit', 'Grocery', 'Shopping', 'Entertainment', 'Insurance', 'Hotels', 'Transportation', 'Airlines'].) I am interested in how much money I will spend in each of the above categories for the first three months of next year.

2. I am getting my data from my monthly transaction history reports from Nov 2019 – Oct 2020. I have created a large data set that concatenated each individual monthly report.

The Apple Card provides monthly reports that looks like this:

	Transaction Date	Clearing Date	Description	Merchant	Category	Type	Amount (USD)
0	04/28/2020	04/30/2020	APPLE.COM/BILL ONE APPLE PARK WAY 866-712-7753...	Apple Services	Other	Purchase	5.35
1	04/28/2020	04/28/2020	MAVERIK #517 24 S GENEVA ROAD VINEYARD 84058 U...	Maverik	Gas	Purchase	10.64
2	04/28/2020	04/28/2020	MICROSOFT*ULTIMATE 1 MONE MICROSOFT WAY MSBILL...	Microsoft	Other	Purchase	16.08

- a. The target is “Amount (USD)”, but this will be divided into each category and different months.
- b. The features are listed above: [Transaction Date, Clearing Date, Description, Category, Type, Amount (USD)].

- c. Yes, I am planning on creating/engineering new features. So far, I have created the features [State, Season, InSchool, Employment] that explain the state the company is located in (categorical), the season of the year (categorical), whether I was in school during that time (Boolean), and whether I was employed during that time (Boolean).
- d. The raw data had dimensions (612, 7). After deleting all rows that were instances where I paid my balance off (since that has no influence of how much money I *spent*), the data has 573 observations.
- e. This data will allow me to answer my question of interest by providing hundreds of rows and over a dozen categories to explain my spending habits. The data will allow me to analyze spending trends by Category and predict new values for each Category across multiple months.