Business Problem and Introduction:

In order to best serve their customers and grow, businesses need to know their customers and their market. By knowing who their customers are, they can target them with specific and relevant products and information that meet their customers' needs, or they can address any potential roadblocks and issues of concern. One method to address this goal is through customer segmentation. Not all customers are the same; they belong to different demographics and showcase different behavioral habits. Segmentation allows us to divide a customer base into groups that are "similar" in some way. These subgroups can help us determine which group of individuals are the most "attractive". How we define similarity and attractiveness depends on the business objective in mind.

In this project, we will focus on a credit card company but hypothetically a similar analysis can be conducted on any customer behavioral data. In this hypothetical case study, our client's marketing and risk department wants to conduct customer segmentation to identify and understand characteristics of each group. Particularly, they want to explore the following questions: what kind of usage behaviors are the customers showing? Are there customers who are more at risk of defaulting? Who uses what service? How frequently do customers make transactions? Which customers are maximizing their credit, and which customers are not?

Ultimately, the company wants you to discover any user behavioral patterns, create user profiles for each segment, and report back to the company's marketing or risk team.

Data:
A credit card company's credit card holders dataset containing behavioral data of users who were active within the last 6 months. Dataset can be found on [Kaggle](#).
- 8950 instances
- 18 attributes: see [Data Dictionary](#)

Anticipated Data Science Approach:
- Clean the data
- Feature engineer additional metrics of Interest such as credit card utilization ratio
- Cluster users on the various metrics created using the unsupervised learning technique, K-Means
- Perform PCA to help reduce dimensions and create visualization of clusters
- Create profiles for each cluster outlining characteristics

Deliverables:
1. Dataset with wrangled metrics (Code included)
2. Code (Juypter notebooks) for each step of data science process
3. Slide deck presentation
4. Documentation