Instead of listing projects , I thought of looking through our course structure and list projects among the data mining topics that will be covered in our class. As such below are my thoughts:

1. Clustering:

Credit card fraud detection: This is a typical classification problem where the task is to classify the abnormal/fraud transactions. This project is popular in Kaggle for classification, but for clustering I guess, it would be pretty decent. Dataset Link: <https://www.kaggle.com/mlg-ulb/creditcardfraud>

Online Retail Customer Segmentation: Here we have to segment customers based on their purchase. Dataset Link : <https://archive.ics.uci.edu/ml/datasets/online+retail>

It would be fine for me to work on clustering of some genetic information as well. (Like clustering the patients, covid variants etc.,) But I am not very sure of the data collection, so if you have any idea of where we could get data, we can have this in our bucket as well).

**Mine**

1. **Text mining - Hate speech recognition in news comments (videos and articles)**
2. **Paper similarity – detecting and counting covid – 19 publications (we can use googlesholar)**
3. **Predicting hereditary disease susceptibility – comparing individual’s genomic data against given GWAS summary statistic data of a given disease. (data available in several sites)**

2. Graph Mining:

This is a new topic which I am not aware of. But I saw this in our course structure. So, adding in a dataset that might be interesting to work with.

<https://paperswithcode.com/dataset/amazon-fraud>