Nicholas Brower
Springboard DSCT - May 2022
Unit 7.1.1 - Capstone Two - Project Ideas

1.   Credit Default Prediction

Use anonymized account activity data to predict credit default among American Express cardholders. This dataset includes purchase, payment, and other account activity information for users within a specific time window.

Data:

https://www.kaggle.com/competitions/amex-default-prediction/data

2.   Book Review Prediction

Predict a book review rating on a scale of 1 to 5 for Goodreads users. This dataset provides information on a user's account activity on Goodreads, including user ID, review dates, review ratings, book identification, and the text of each review. Build a model to predict a user's score given their history, the text of a review, and the ID of a book.

Data:

https://www.kaggle.com/competitions/goodreads-books-reviews-290312/data

3.   Pennsylvania Roadway Maintenance

Given roadway condition, traffic volume, construction, weather, and maintenance data, predict when segments of roadways will require maintenance. Either focus on Pennsylvania's most traveled road segments, or limit data to road segments in single county.

There are many datasets relevant to the creation of a predictive model. Pennsylvania's Roadway Mangement System and Department of Transportation websites offer useful introductions to much of the required roadway data. Weather data is easily obtained. Underlying soil compositions may be relevant for some roads; the US Department of Agriculture offers this data in many useful formats.

Info:

PA Roadway Mangement System
https://www.penndot.pa.gov/ProjectAndPrograms/ResearchandTesting/RoadwayManagementandTesting/Pages/Road-Management-System.aspx

PennDOT Open Data
https://data-pennshare.opendata.arcgis.com/

Data:

Traffic Volumes
https://data-pennshare.opendata.arcgis.com/datasets/PennShare::rmstraffic-traffic-volumes/about

Transportation Improvement Projects – Points
https://data-pennshare.opendata.arcgis.com/datasets/PennShare::transportation-improvement-projects-points/about

Transportaiton Improvement Projects - Lines
https://data-pennshare.opendata.arcgis.com/datasets/PennShare::transportation-improvement-projects-lines/about

Weather:
http://www.climate.psu.edu/data/city_information/index.php?city=phl&page=dwa&type=big7

USDA Soil Survey
https://www.nrcs.usda.gov/wps/portal/nrcs/surveylist/soils/survey/state/?stateId=PA