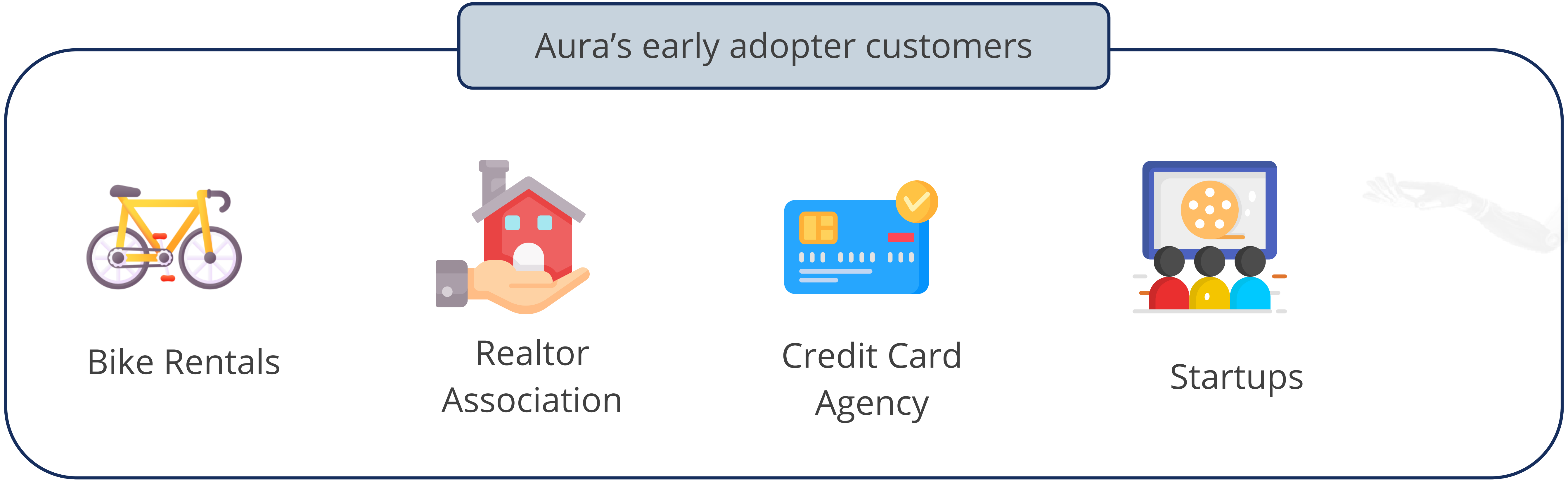
DATA AND
ARTIFICIAL INTELLIGENCE

Capstone Session 8

Machine Learning for Modeling

Machine Learning End Goal

The intelligence provided by Aura will help customers make decisions for their omnichannel marketing and customer acquisition programs.



Aura’s early adopter customers

- Bike Rentals
- Realtor Association
- Credit Card Agency
- Startups

Project Statement

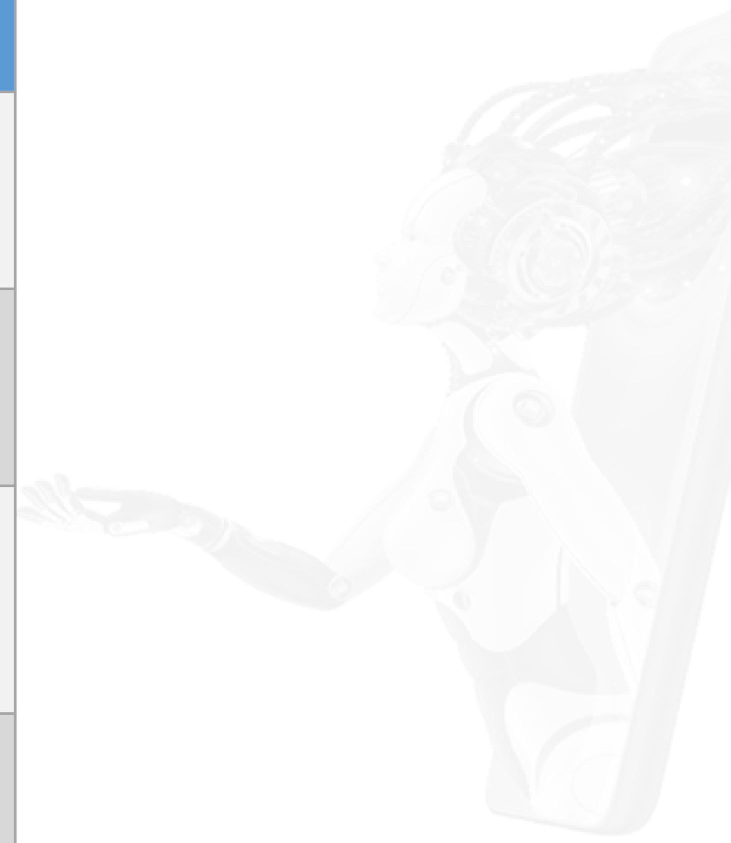
Aura must do the following:

- Predict bike-sharing demand
- Classify incomes
- Cluster credit card users
- Build a recommendation engine

Week 8: Dataset Description

ratings.csv

Variable	Description
userId	A unique Id represents every user.
movieId	Unique ID representing a movie
rating	Represents the rating given by the user to the corresponding movie.
timestamp	The time at which the rating was recorded.



Week 8: Dataset Description

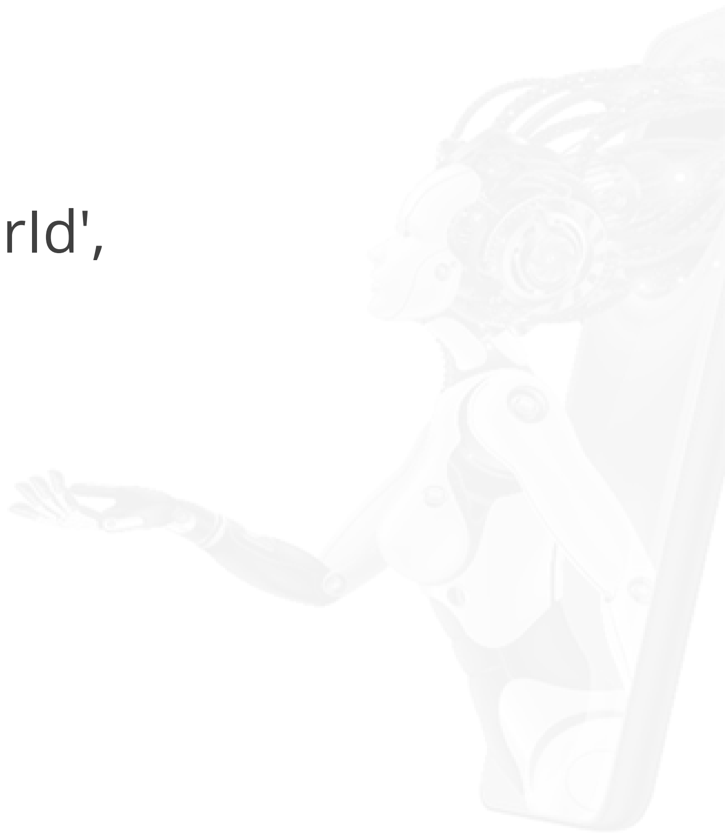
movies.csv

Variable	Description
movieId	A unique Id represents every user.
title	Movie name which is represented by the corresponding movieId
genres	Represents the category of the movie.
timestamp	The time at which the rating was recorded.

Week 8

Task: Study the various Recommendation Techniques for recommending movies using movies.csv, ratings.csv datasets

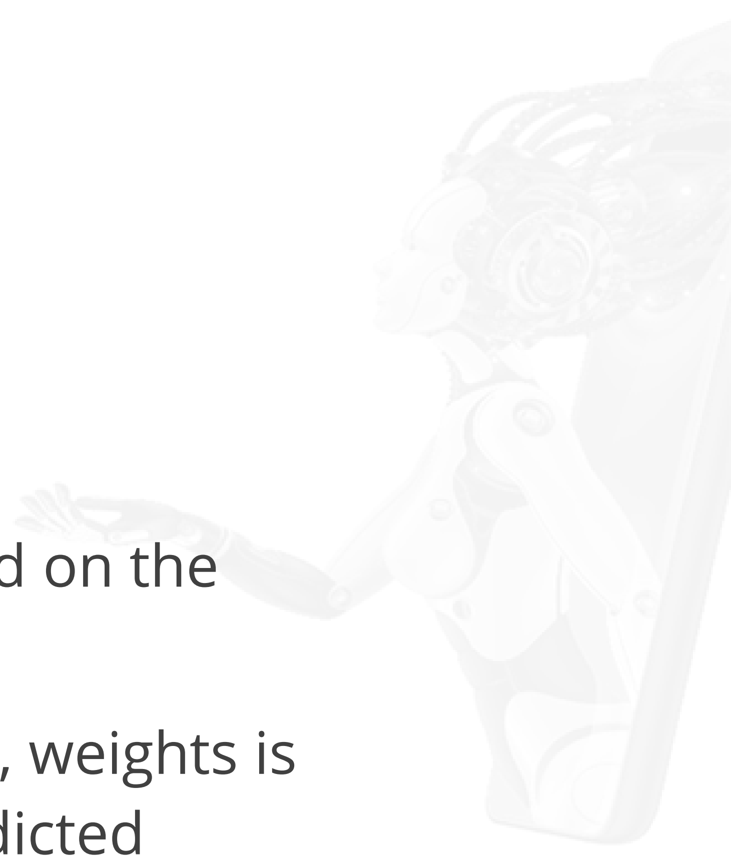
- Load movies.csv and ratings.csv dataset
- Merge both data frames on movieid
- Create User-Item Matrix (Hint: Use pandas pivot_table method with index = 'userId', columns = 'title', values = 'rating')



Week 8

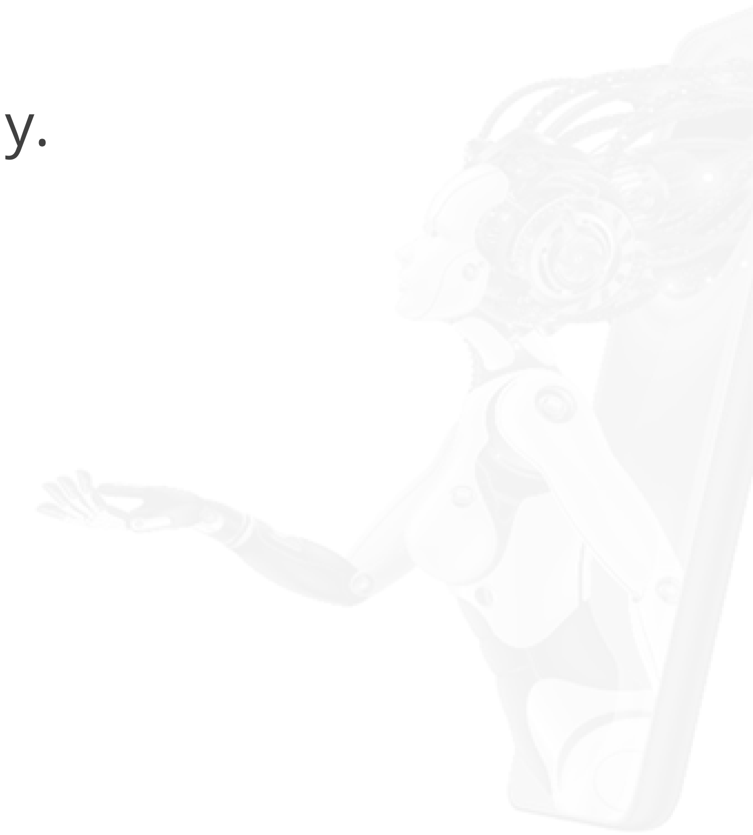
- Perform User-based Collaborative Filtering
  - Fill the row-wise NaNs in the User-Item Matrix with the corresponding user's mean ratings, and find the Pearson correlation between users
  - Choose the correlation of all users with User 1 only.
  - Sort the user 1 correlation in the descending order
  - Drop the NaN values generated in the correlation matrix
  - Choose the top 50 users that are highly correlated to User 1
  - Predict the rating that User 1 might give for the movie with movieid 32 based on the top 50 user correlation matrix

  (Hint: Predicted rating = sum of [(weights) * (ratings)] / sum of (weights). Here, weights is the correlation of the corresponding user with the first user.). That is, the predicted rating is calculated as the weighted average of k similar users



Week 8

- Perform Item-based Collaborative Filtering
  - Fill the column-wise NaN's in the User-Item Matrix with the corresponding movie's mean ratings, and find Pearson correlation between movies
  - Choose the correlation of all movies with the movie Jurassic Park (1993) only.
  - Sort the Jurassic Park movie correlation in descending order
  - Drop the NaN values generated in the correlation matrix
  - Find the most 10 movies similar to the movie Jurassic Park (1993)



Week 8

- Perform KNNBasic, SVD, NMF Model-based Collaborative Filtering
  - Initialize KNNBasic with similarity configuration as Mean Squared Distance Similarity (msd), 20 neighbors and cross-validate 5 folds against measure RMSE.

  (Hint: cross_validate(algo=algo, data=data, measures=['RMSE'], cv=5, verbose=True))

  - Initialize Singular Value Decomposition (SVD) and cross-validate 5 folds against measure RMSE.
  - Initialize Non-Negative Matrix Factorization (NMF)and cross-validate 5 folds against measure RMSE.
  - Print best score and best params from Cross Validate on all the models built.

Thank You