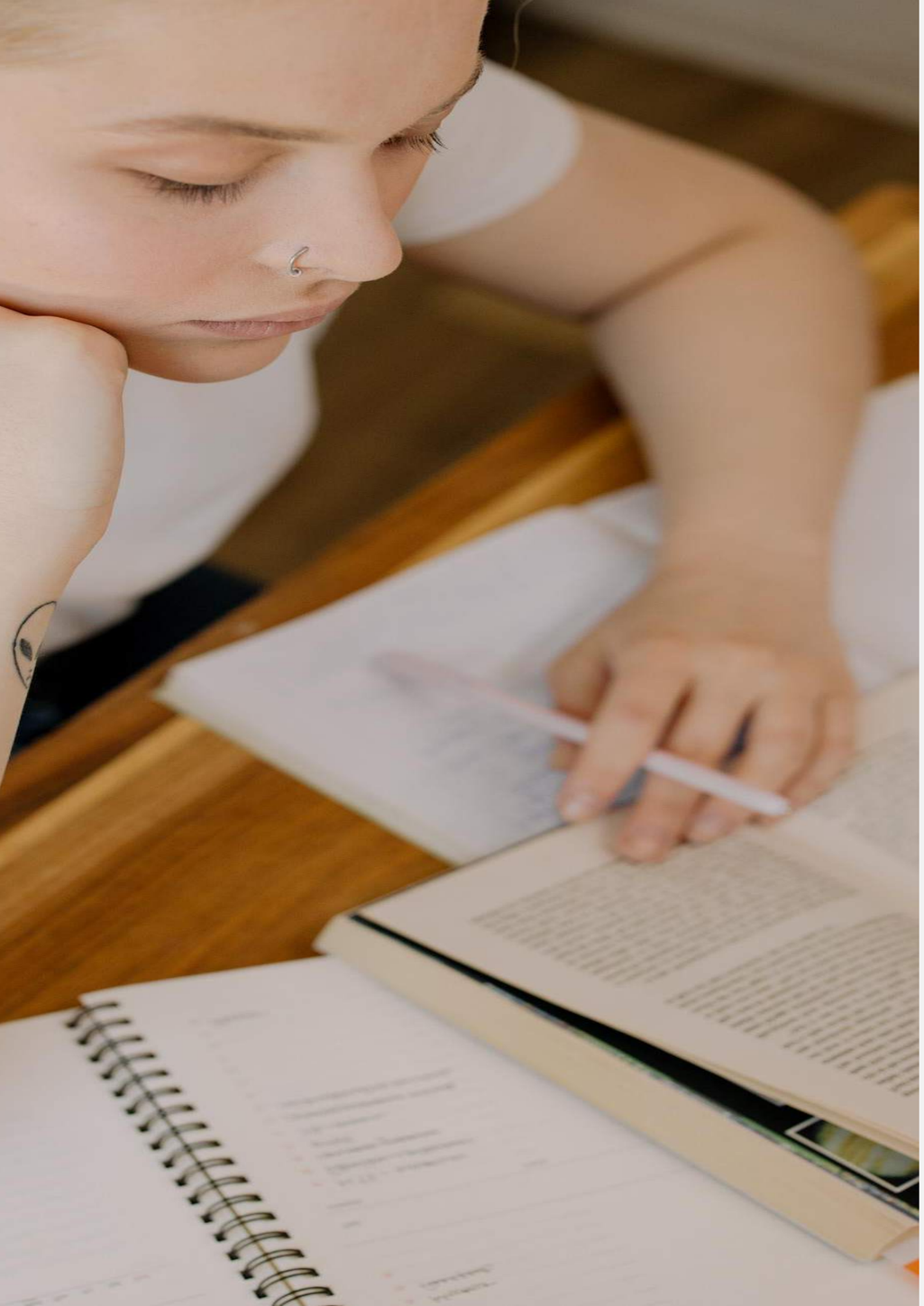
COURSE RECOMMENDATION SYSTEM

Data Wranglers

George Mason University

Under the guidance of

Dr. Isaac K Gang



AGENDA

- Introduction
- Problem Statement
- Data Preprocessing and Analysis
- Visualizations
- Recommendation Systems
- Model Evaluation
- Demonstration
- Future work
- Q&A

OUR TEAM

Girija Mangrulkar

Scrum Master

Abhishek Godavarthi

Product Owner

Jaswanth Kunku

Developer

Srikanth Konne

Developer

Sanket Modak

Developer

Vishveshwar Kondala

Developer

MEET OUR CLIENT

BRAINY n BRIGHT

- IDEATE. INNOVATE. CREATE.

ACCREDITATION

STEM.ORG ACCREDITED EDUCATIONAL EXPERIENCE

Microsoft Qualified Academic Institution

حكومة دبي GOVERNMENT OF DUBAI DUBAI Knowledge المعرفة

ACTVET Abu Dhabi Centre for Technical and Vocational Education and Training

Aamir Rizwan , CEO

About the Client:

- Center of Science & Technology based in Dubai.
- Provides training services to KG to 12th grade students in 30 nationalities
- Offers competitions and internships on 21st century STEM courses
- Awarded as the Best Global STEM Institution 2021 & 2022

Brainy n Bright - Best Robotics Courses & STEM Training Center in Dubai. (n.d.). https://brainynbright.com/

Problem Statement

In today's fast-paced and competitive world, people and businesses want an effective course selection system to help them meet their learning and development objectives. Traditional course recommendation algorithms often struggle to effectively forecast the courses that would be of interest to an individual or group. Our goal is to develop a collaborative recommendation system that provides recommendations to students based on the similar interests.

Domain of problem: Education, E-Learning

Features of the proposed Application:

- Gathers student information based on input and ingests it to the model.
- Recommends highly rated courses based on similar students in the same age group.
- Synchronizes with the database to generate recommendations for new users.

Dataset

Data Type: Proprietary (.numbers format – converted to CSV)

Data Properties: 12 columns with 9806 records (~1.3 MB)

Student Name	Course Selected	Nationality	School	Demographic Data	Age	Modes of Learning	Session	Field of Interest	Difficulty Level	Course Recommended	Course Rating
student 1	Robotics Innovative Learning	UK	Suchitran	Female	10	Online	24	Art	Applied Tech	Robotics Innovative Learning	5
student 2	Robotics Innovative Learning	Poland	Indian School Bahrain	Male	10	Online	96	Mathematics	Beginner	Robotics Innovative Learning	5
student 3	Spring Camps	China	Indian School Muscat, Oman	Male	10	In-person	48	Music	Proficiency	Python	1

Data Quality:

- **Accuracy** : ✅
- **Integrity** : ✅
- **Conformity** : ✅
- **Consistency** : ❌

Tools and Architecture

- **Programming and Databases :** Python, MySQL
- **Modeling Environment :** Jupyter Notebook, Visual Studio Code , Anaconda
- **Dynamic Visualizations :** Tableau

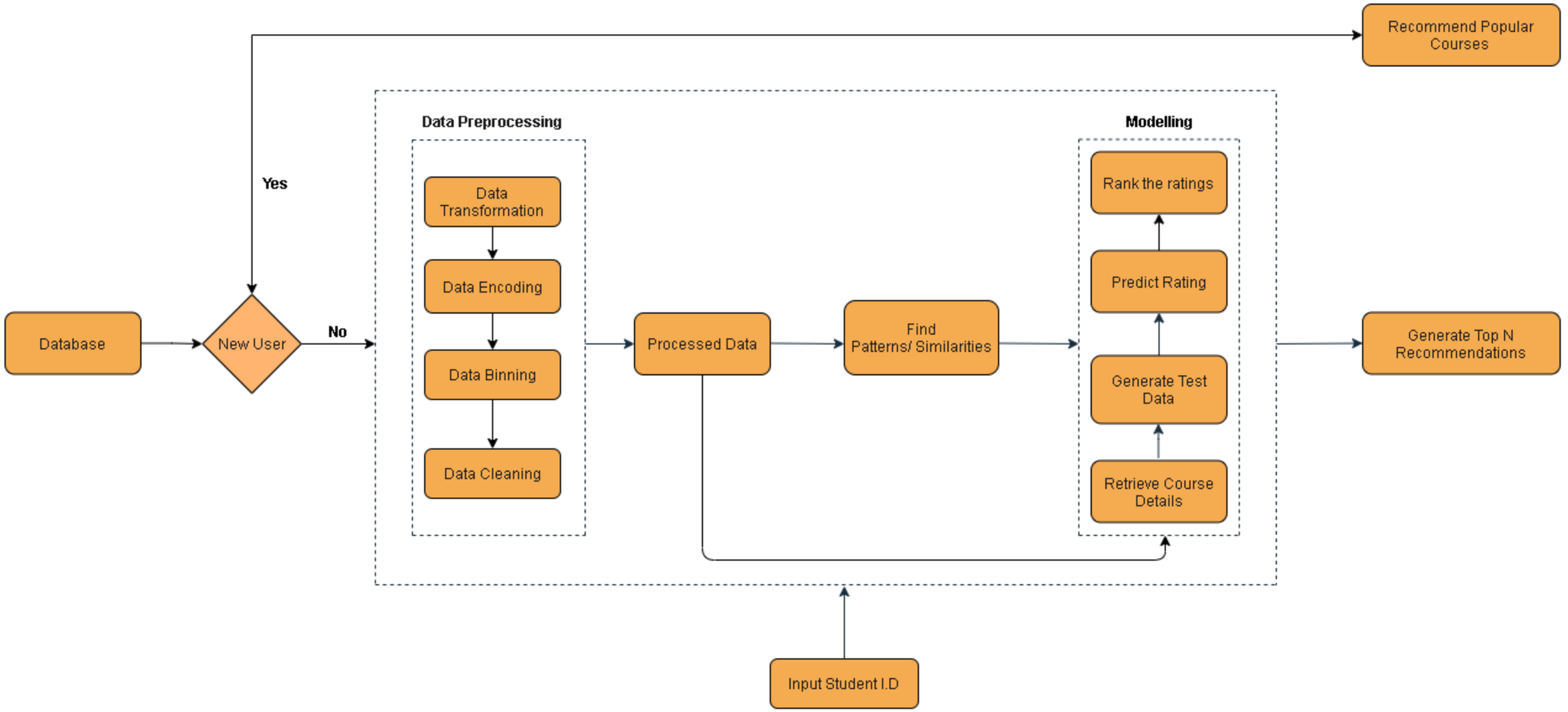


Image produced using draw.io

Data Preprocessing and Analysis

Data Formatting of Columns

Data Cleaning In courses

Data Encoding and Filtering on Age

Generated fields based on age factor

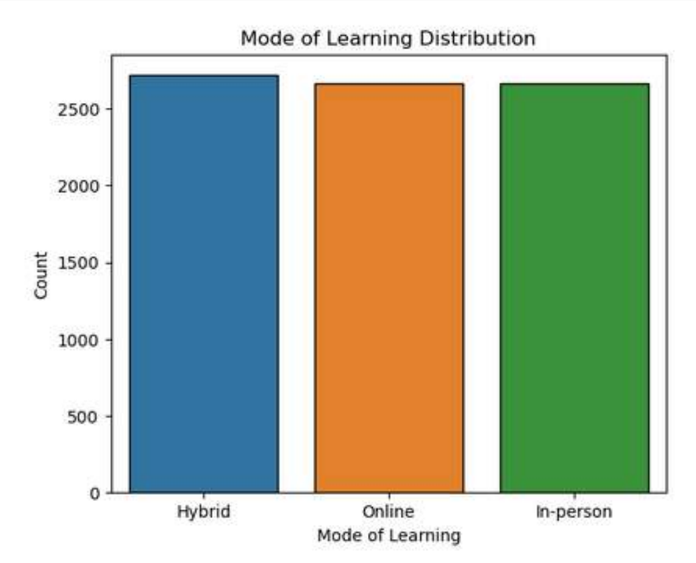
Unique Identifier for student and course names

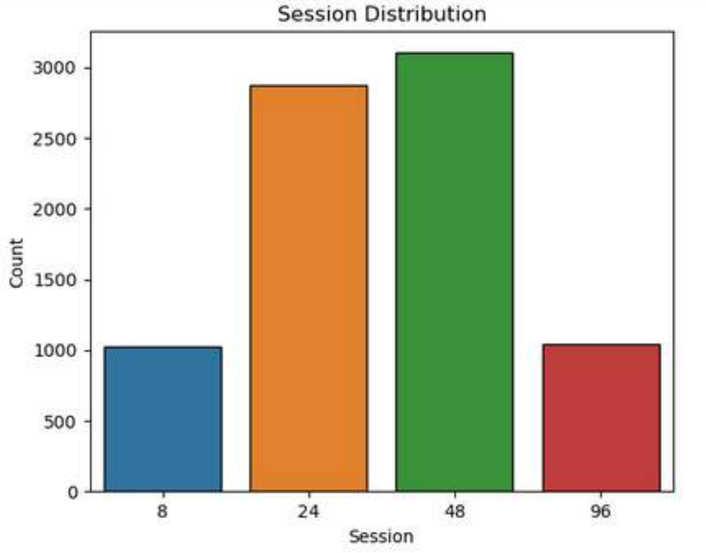
Final Dataset:

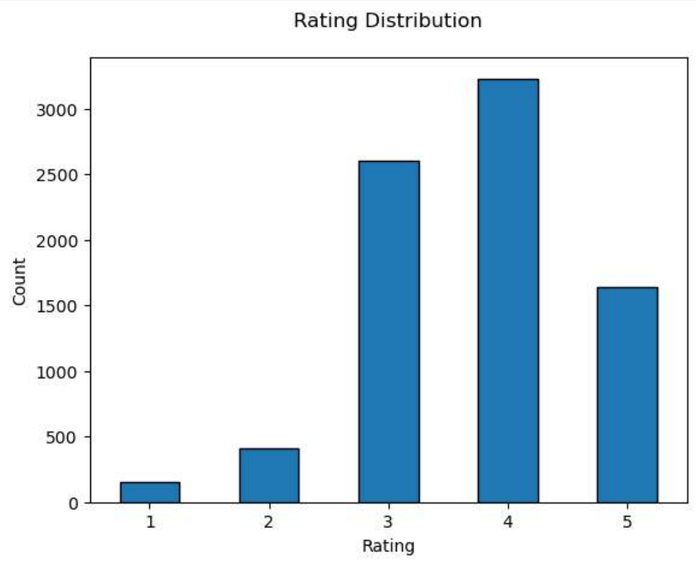
- No of records dropped by 17 %
- No of Courses reduced from 109 to 45
- Generated 3000 student id’s

Student Enrollment	Percentage
One Course	19.9
Two Courses	26.1
Three Courses	24.2
More than 3	29.8

Data Analysis







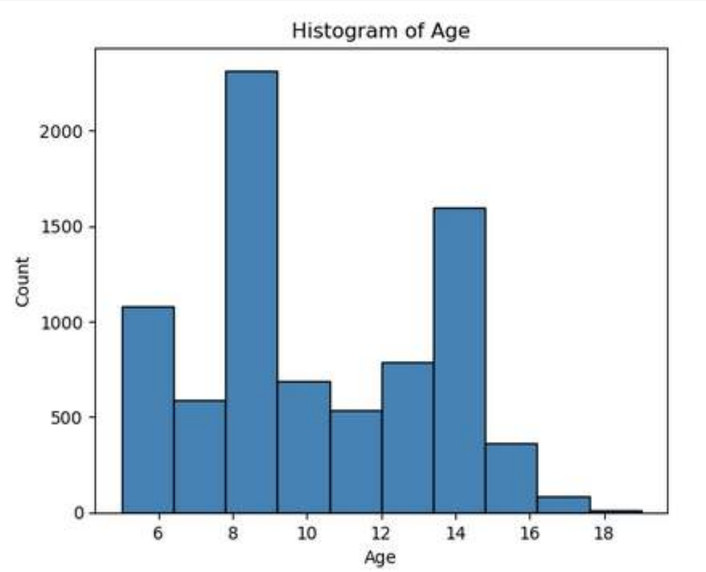


Image Source : Python (matplotlib)

Visualizations

Image Source : Tableau

Visualizations

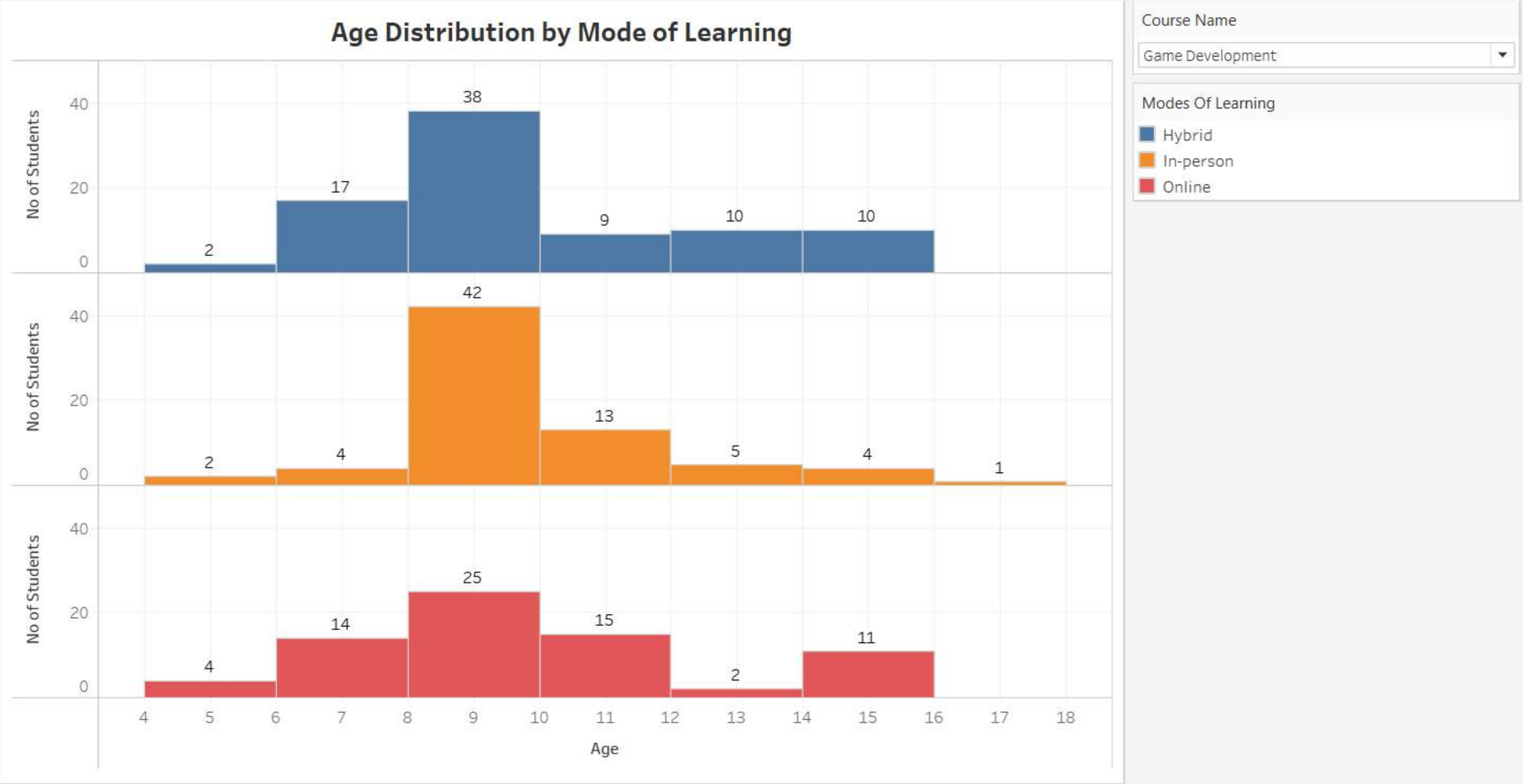


Image Source : Tableau

Visualizations

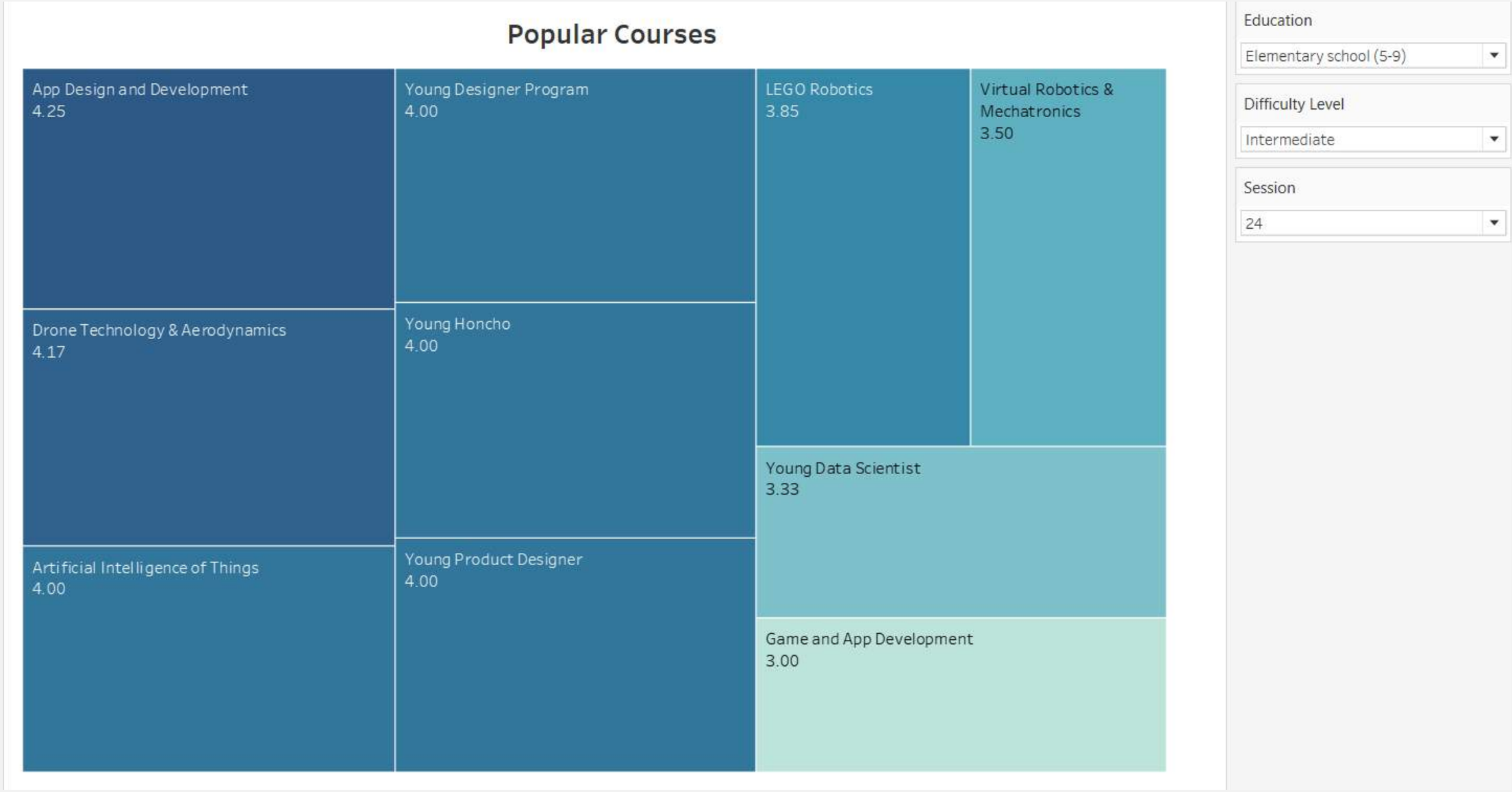


Image Source : Tableau

Visualizations

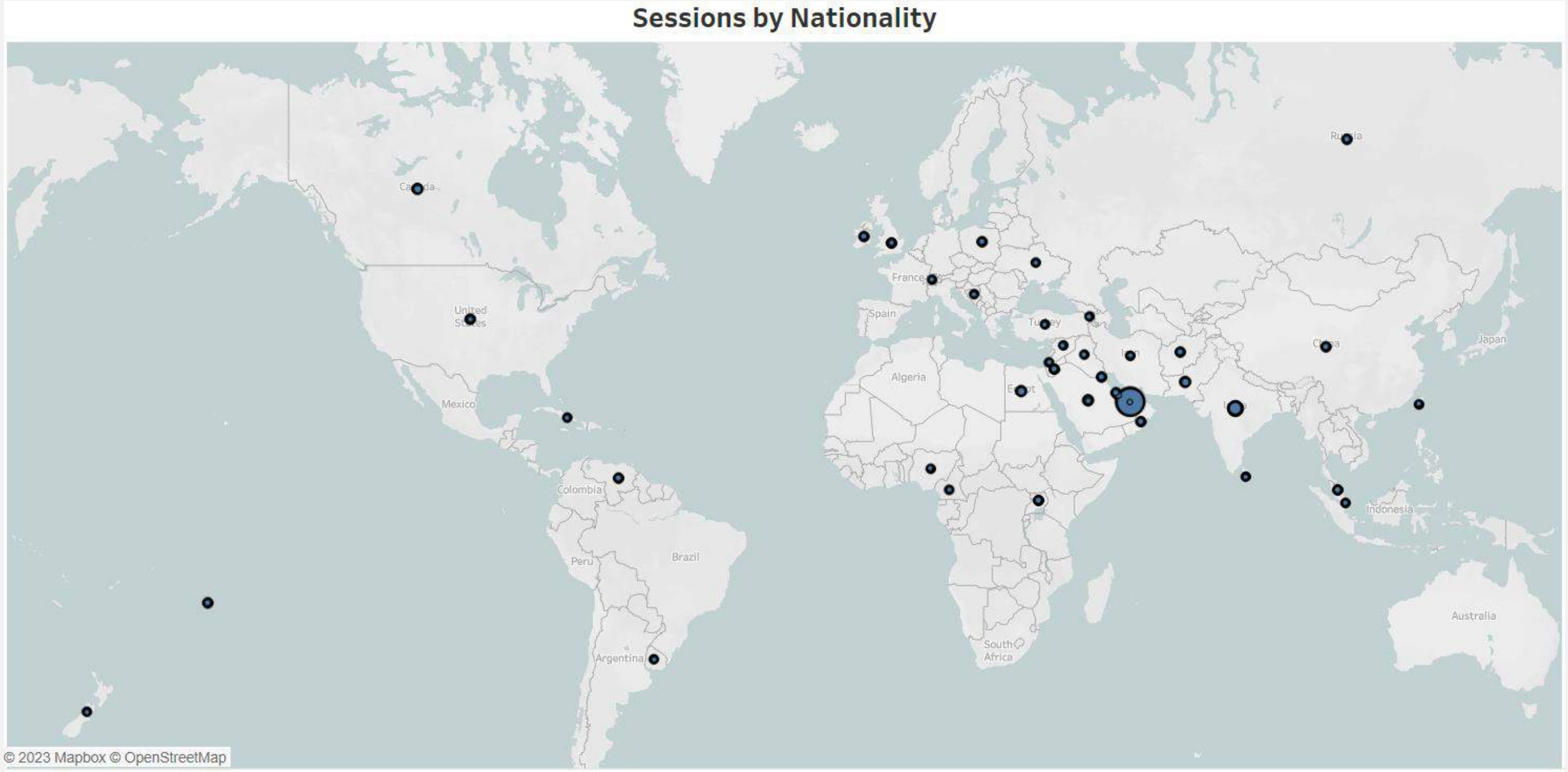
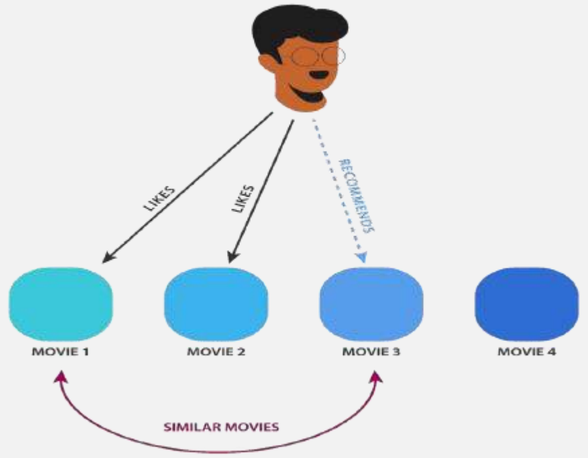


Image Source : Tableau

Recommendation Systems

Content Based :

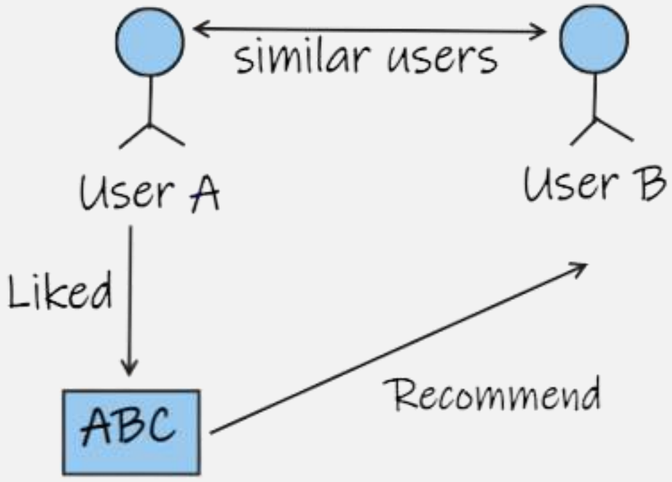
- Suggests items to users based on their preferences, interests, and past behavior.
- Recommending courses based on their features is only possible if the course has a distinct set of features and a list of the user's options.
- Creates user profile based on the preferences.



Collaborative Based :

- In this approach the characteristics of each individual course are not considered.
- To suggest new products, collaborative filtering-based recommender systems rely primarily on previous interactions between users and courses
- The history data of the user interacting with the courses is recorded and saved in collaborative filtering.

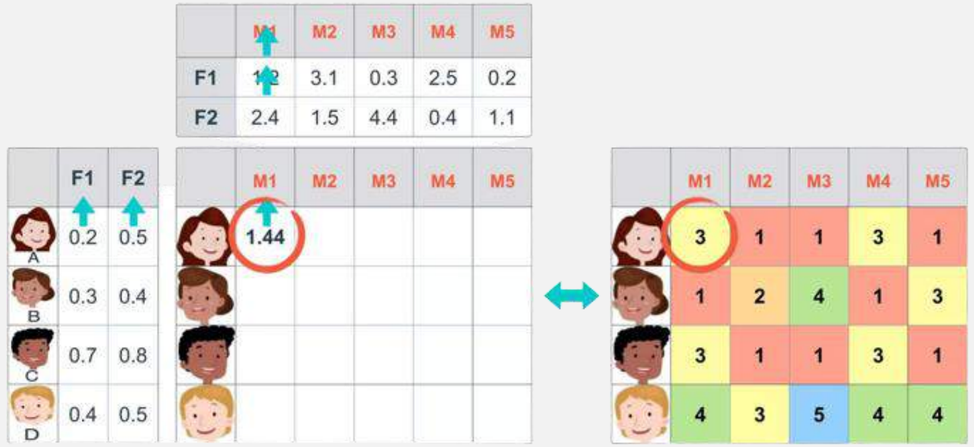
Distanced Based Algorithm: K-Nearest Neighbors (KNN)



Collaborative Filtering - Matrix Factorization

How it works:

1. Decompose the user-item matrix into smaller matrices
   a) User Factor Matrix (U)
   b) Item Factor Matrix (V)
2. Generate Latent Factors
3. Estimate the Bias using
   - Alternative Least Squares methods (ALS)
4. Compute the rating using the dot product of the latent factors from U and V



Algorithms:

- **SVD (Singular Value Decomposition)** : Baseline Matrix Factorization Algorithm with ALS
- **SVD++ :** Extension to SVD with implicit feedback
- **NMF(Non-Negative Matrix Factorization):** SVD extension with positive user and item factors

Advantages:

- Accuracy
- Scalability
- No Cold-start

Model Evaluation

Data Input : StudentID, Course ID, Course Rating *(actual)*

Data Output : Course Rating *(predicted)*

Model	Test RMSE	Train Time	Test Time
SVD	0.854	0.022	0.004
SVDpp	0.857	0.021	0.006
KNN	0.902	0.111	0.043
NMF	1.102	0.092	0.003

Synthetic Data :

- **Package** : Synthetic Data Vault (SDV)
- **Model :** Fast ML
- **Dataset Information:** 50000 records with 35000 users

Observations :

- **Distance Based :** Out of Memory, Increased Train/Test times
- **Matrix Factorization :** Fast & Significant decrease in RMSE

Hyper parameter Tuning :

1. **Best Parameters :** factors=100, epochs=5, learning rate=0.005, regularization=0.1
2. **Best RMSE :** 0.84

Model Evaluation

$$P = \frac{\#\text{ of our recommendations that are relevant}}{\#\text{ of items we recommended}}$$

Precision

$$r = \frac{\#\text{ of our recommendations that are relevant}}{\#\text{ of all the possible relevant items}}$$

Recall

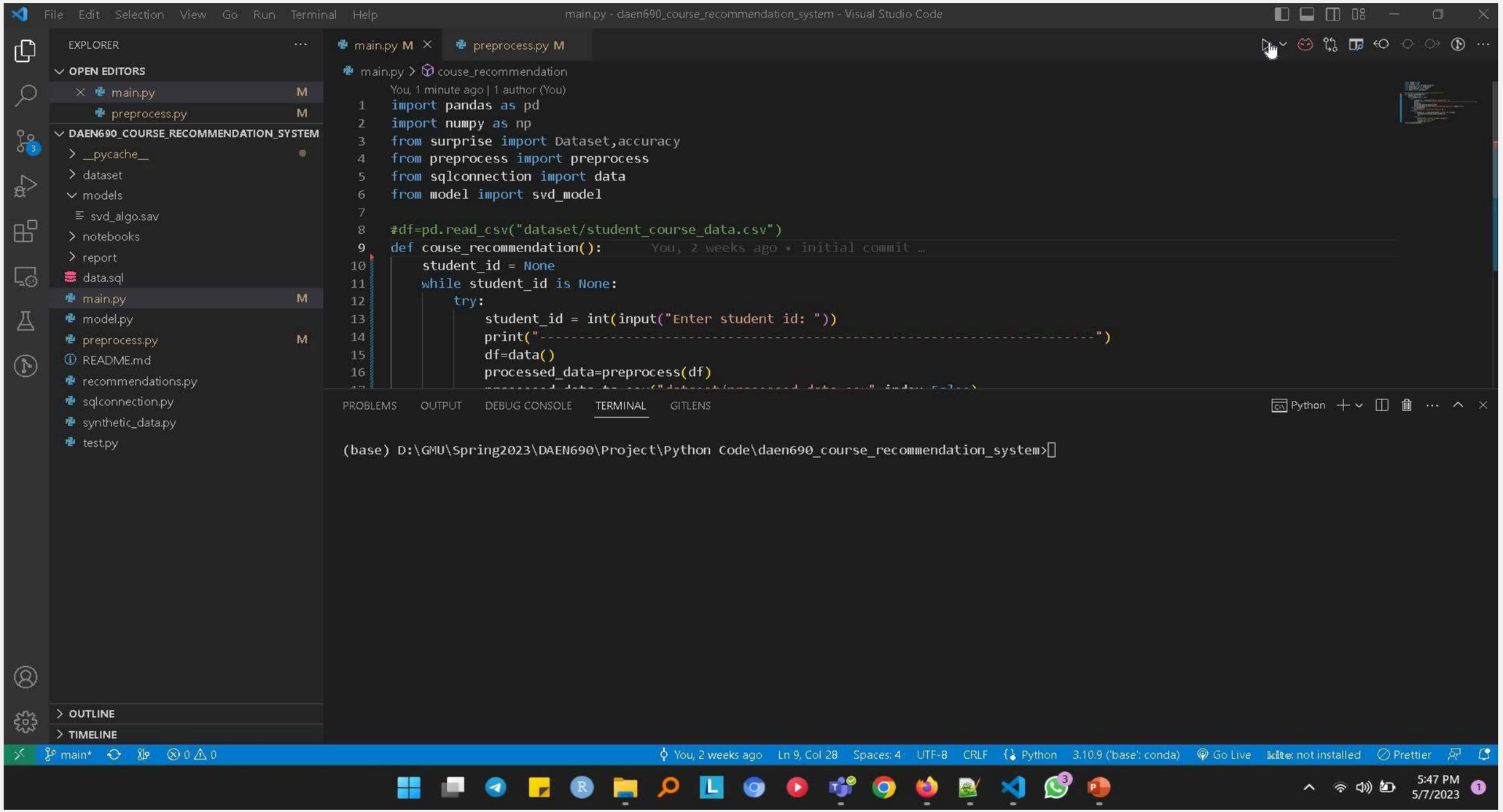
Confusion Matrix

	Recommended	**Not Recommended**
Relevant	True Positive	False Negative
Not Relevant	False Positive	True Negative

Parameters : K=5 , Relevance Threshold = 3.5

Model	**Mean Average Precision @ k**	**Mean Average Recall @ k**
SVD	0.70	0.82
SVDpp	0.70	0.82
NMF	0.67	0.87
KNN	0.63	0.93

Demonstration



GitHub Repository : https://github.com/jaswanth333/daen690_course_recommendation_system

FUTURE WORK

01 Incorporating additional data fields such as course description,domain,reviews

02 Implement Deep learning models or hybrid models.

03 Gain Domain knowledge on social and contextual elements that influence course selection

THANK YOU