**Project Based Learning Report**

on

**Design and Build Movies Recommendation System Project**

Submitted in the partial fulfillment of the requirements

For the Project based learning in (**Essentials of Data Science**)

in

Electronics & Communication Engineering

By

**2014111129 Arya Tripathi**

**2014111112 Ankush Singh**

**2014111140 Geetika Shaaran**

Under the guidance of Course In-charge

Prof. Dnyanesh S.Lavhkare

Department of Electronics & Communication Engineering

Bharati Vidyapeeth

(Deemed to be University)

College of Engineering,

Pune – 4110043

**Academic Year: 2021-22**

**Bharati Vidyapeeth**

**(Deemed to be University)**

**College of Engineering,**

**Pune – 411043**

**DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING**

**CERTIFICATE**

Certified that the Project Based Learning report entitled, **Design and Build Movies Recommendation System Project**

is work done by

**2014111129 Arya Tripathi**

**2014111112 Ankush Singh**

**2014111140 Geetika Shaaran**

in partial fulfillment of the requirements for the award of credits for Project Based Learning (PBL) in **Essentials of Data Science Course** of Bachelor of Technology Semester IV, Electronics & Communication Engineering.

**Date: 21 May 2022**

**Prof. Dnyanesh S.Lavhkare Dr. Tanuja S.Dhope**

**Course In-charge PBL Co-Ordinator**

**Dr. Arundhati A.Shinde**

**Professor & Head**

**ELECTRONICS & COMMUNICATION ENGINEERING**

|  |  |
| --- | --- |
| **Index: -** | |
| Page No. | Contents |
| 1-1 | Problem Statement with Solution |
| 2-5 | Description about project |
| 6-9 | Software Used |
| 10-13 | Results with Analysis |
| 14-14 | Conclusion & Outcome |

**Problem Statement :-**

What is Data Science? Why learn Data Science?

**Solution :-**

Data science is the domain of study that deals with vast volumes of data using modern tools and techniques to find unseen patterns, derive meaningful information, and make business decisions. Data science uses complex machine learning algorithms to build predictive models. The data used for analysis can come from many different sources and presented in various formats.

Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data. Data science practitioners apply [machine learning](https://www.datarobot.com/wiki/machine-learning/) [algorithms](https://www.datarobot.com/wiki/algorithm/) to numbers, text, images, video, audio, and more to produce [artificial intelligence (AI)](https://www.datarobot.com/wiki/artificial-intelligence/) systems to perform tasks that ordinarily require human intelligence. In turn, these systems generate [insights](https://www.datarobot.com/wiki/insights/) which analysts and business users can translate into tangible business value.

Reasons to learn Data Science are: -

1. Learning about data science provides an opportunity for you to recreate yourself.
2. **We live in a digital world, everything is data-driven.** There is data science in **business, accounting, education, science, engineering, healthcare, technology, energy sector, government**, and so on.
3. **Data science is also a very promising field with lots of high paying job opportunities.**
4. **Basic data science skills are important for personal use.**
5. Great potential to branch out with different options.
6. Become a decision-maker, not every job opportunity will give you the power to make informed business decisions. For a data scientist, that is the core responsibility.
7. Less competitive because it is a highly analytical role, competition is less, but demand is not. With a limited talent pool, there is always a challenge for businesses to hire in these roles.

**1**

**How to Build a Movie Recommendation System**

Step by step guide to building a simple recommendation system

Have you ever wondered how YouTube recommends content, or how Facebook recommends you, new friends? Perhaps you’ve noticed similar recommendations with LinkedIn connections, or how Amazon will recommend similar products while you’re browsing. All of these recommendations are made possible by the implementation of recommender systems.

Recommender systems encompass a class of techniques and algorithms that can suggest “relevant” items to users. They predict future behavior based on past data through a multitude of techniques including matrix factorization.

In this article, I’ll look at why we need recommender systems and the different types of users online. Then, I’ll show you how to build your own movie recommendation system using an open-source dataset.

**Contents**

* Why Do We Need Recommender Systems?
* Types of Recommender Systems  
  A) Content-Based Movie Recommendation Systems  
  B) Collaborative Filtering Movie Recommendation Systems
* The Dataset
* Designing a Movie Recommendation System
* Implementation  
  Step 1: Matrix Factorization-based Algorithm  
  Step 2: Creating Handcrafted Features  
  Step 3: Creating a Final Model for our Movie Recommendation System
* Performance Metrics
* Summary



**Why Do We Need Recommender Systems?**

We now live in what some call the “era of abundance”. For any given product, there are sometimes thousands of options to choose from. Think of the examples above: streaming videos, social networking, online shopping; the list goes on. Recommender systems help to personalize a platform and help the user find something they like.

The easiest and simplest way to do this is to recommend the most popular items. However, to really enhance the user experience through personalized recommendations, we need dedicated recommender systems.

From a business standpoint, the more relevant products a user finds on the platform, the higher their engagement. This often results in increased revenue for the platform itself. Various sources say that as much as 35–40% of tech giants’ revenue comes from recommendations alone.

Now that we understand the importance of recommender systems, let’s have a look at types of recommendation systems, then build our own with open-sourced data!



**Types of Recommender Systems**

Machine learning algorithms in recommender systems typically fit into two categories: content-based systems and collaborative filtering systems. Modern recommender systems combine both approaches.

Let’s have a look at how they work using movie recommendation systems as a base.

**A) Content-Based Movie Recommendation Systems**

Content-based methods are based on the similarity of movie attributes. Using this type of recommender system, if a user watches one movie, similar movies are recommended. For example, if a user watches a comedy movie starring Adam Sandler, the system will recommend them movies in the same genre or starring the same actor, or both. With this in mind, the input for building a content-based recommender system is movie attributes.

Figure 1: Overview of content-based recommendation system (Image created by author)

**B) Collaborative Filtering Movie Recommendation Systems**

With collaborative filtering, the system is based on past interactions between users and movies. With this in mind, the input for a collaborative filtering system is made up of past data of user interactions with the movies they watch.

For example, if user A watches M1, M2, and M3, and user B watches M1, M3, M4, we recommend M1 and M3 to a similar user C. You can see how this looks in the figure below for clearer reference.

**Software Used: -**

**Google colab-**

Google Colab was developed by Google to provide free access to GPU’s and TPU’s to anyone who needs them to build a machine learning or deep learning model. Google Colab can be defined as an improved version of Jupyter Notebook..

Programming Languages are an intermediate form between human-understandable language and machine understandable language. Every application is built using one of the many programming languages available. Maybe a person with a computer science background can understand, but not everyone can. Remember, as Software Developers, we develop applications for people with little computer science knowledge.

Consider you are creating a machine learning model to improve customer satisfaction for a local store, in that case you will have to explain how the model can do this task, and you can’t just explain him with your code base. Most people facing this situation will prepare a separate presentation. Notebooks were created so that it is not necessary. Notebook documents can include executable lines of code along with text, images, figures, tables, graphs, equations, and much more graphical data. In simple words, Notebook documents are a way of creating human-readable executable documents.

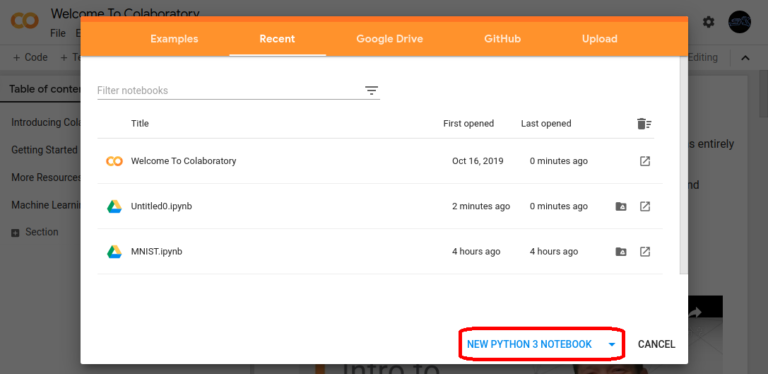
## Google Colab Features

Google Colab provides tons of exciting features that any modern IDE offers, and much more. Some of the most exciting features are listed below.

* Interactive tutorials to learn machine learning and neural networks.
* Write and execute Python 3 code without having a local setup Execute terminal commands from the Notebook.



[This Photo](https://www.prometheus-studio.it/prometheus_blog_wp/2019/09/10/colab-il-jupyter-notebook-di-google/) by Unknown Author is licensed under [CC BY-SA-NC](https://creativecommons.org/licenses/by-nc-sa/3.0/)



[This Photo](https://blog.toright.com/posts/6725/google-colab-free-gpu-ai-train.html) by Unknown Author is licensed under [CC BY-ND](https://creativecommons.org/licenses/by-nd/3.0/)

* Import datasets from external sources such as Kaggle.
* Save your Notebooks to Google Drive.
* Import Notebooks from Google Drive.
* Free cloud service, GPUs and TP
* Integrate with PyTorch, Tensor Flow, Open CV.
* Import or publish directly from/to GitHub.

## How to Use Google Colab?

Just like any other product from Google, you need a Google account to get started. You can access Google Colab using this link Welcome To Colaboratory – Colaboratory (google.com) and signing in through your Google account.

You will land on a page similar to the one shown below.

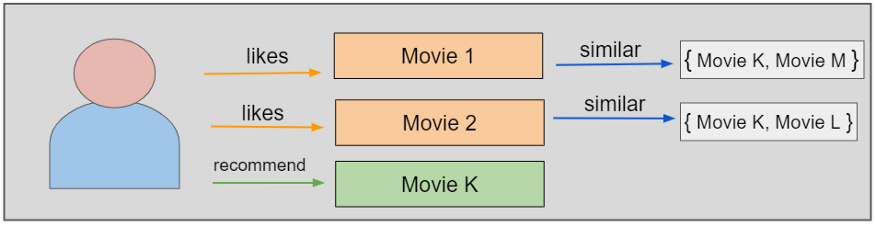
**5**

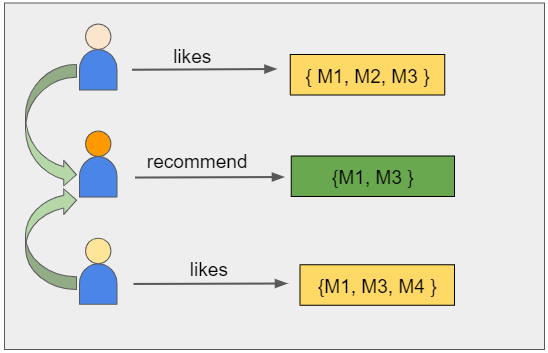
**Result with Analysis**

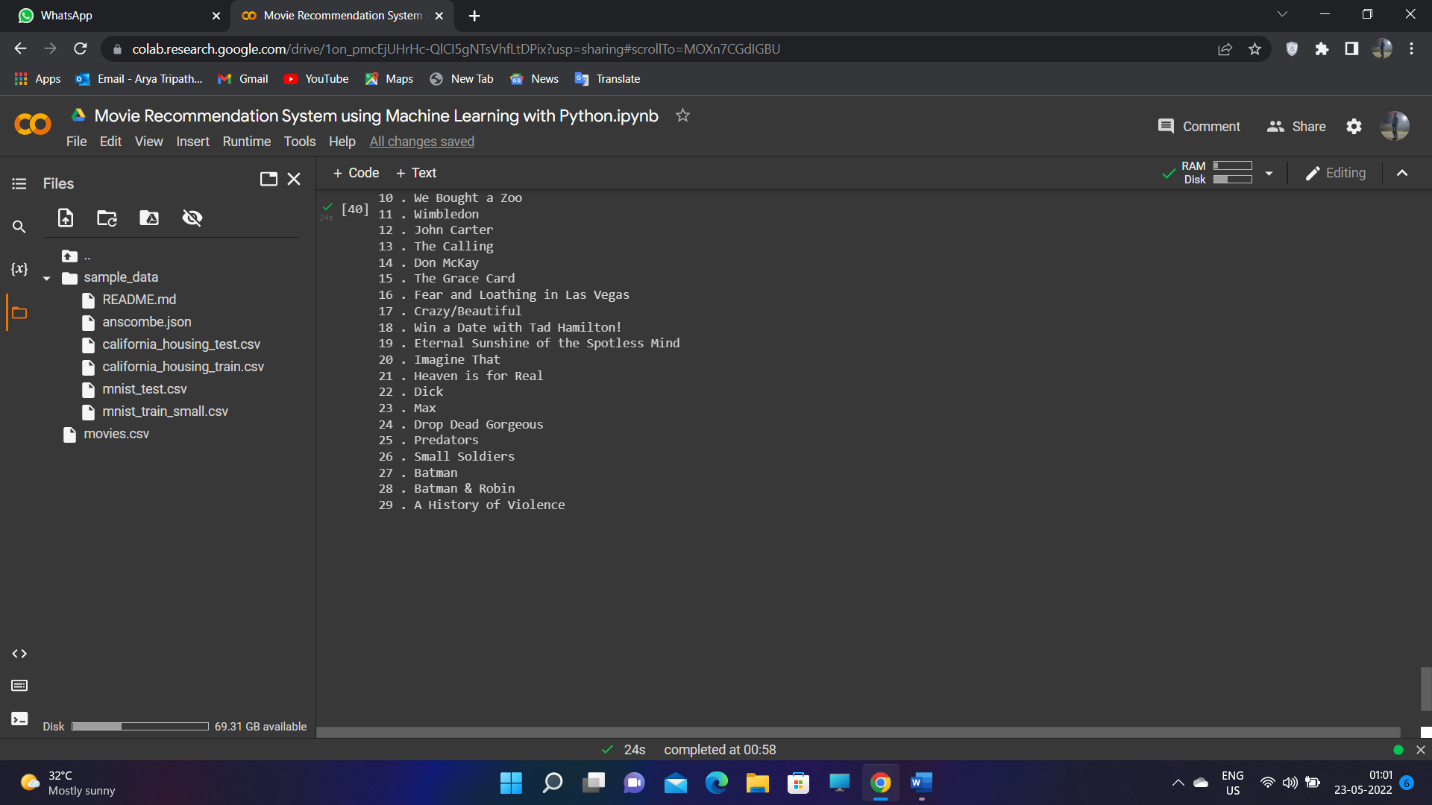
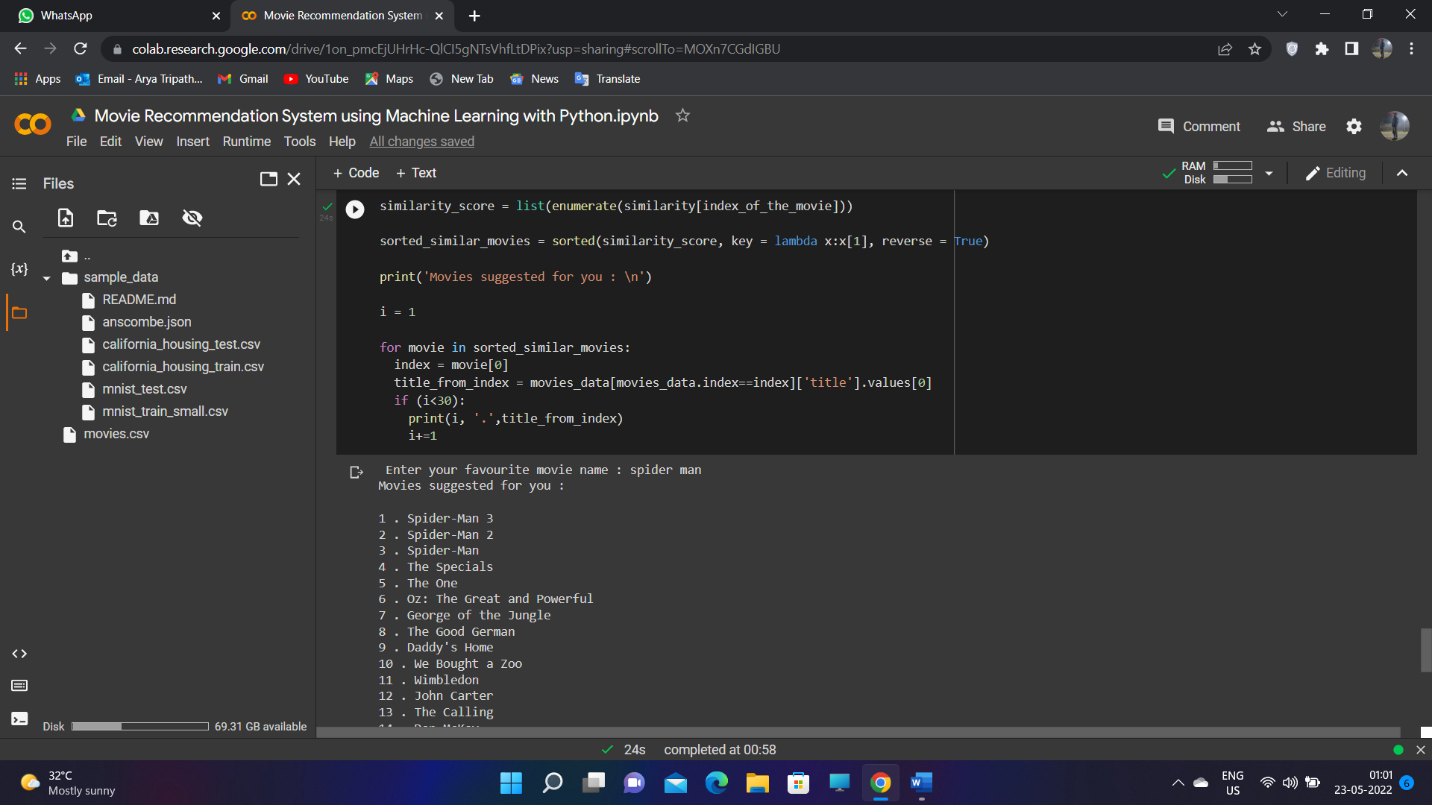
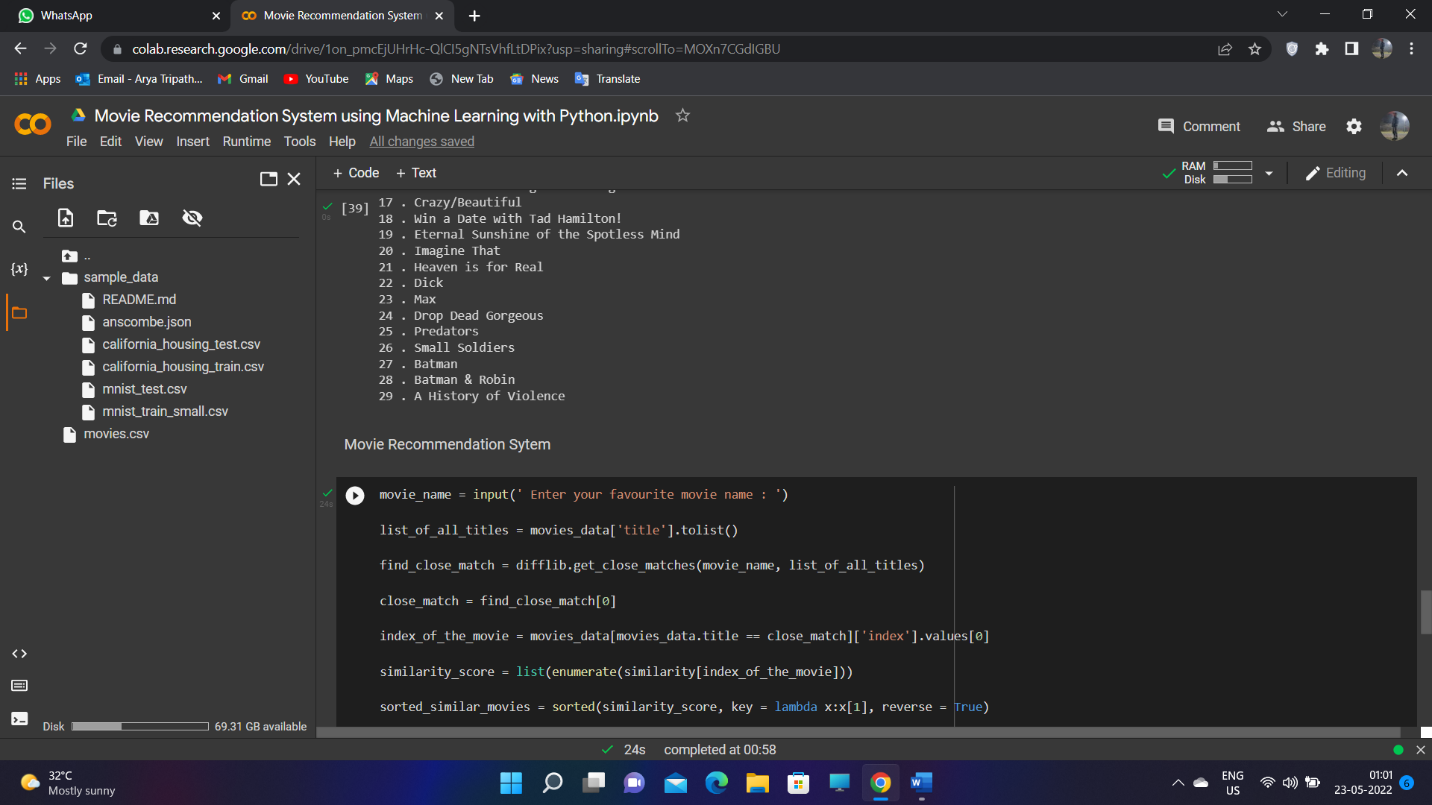
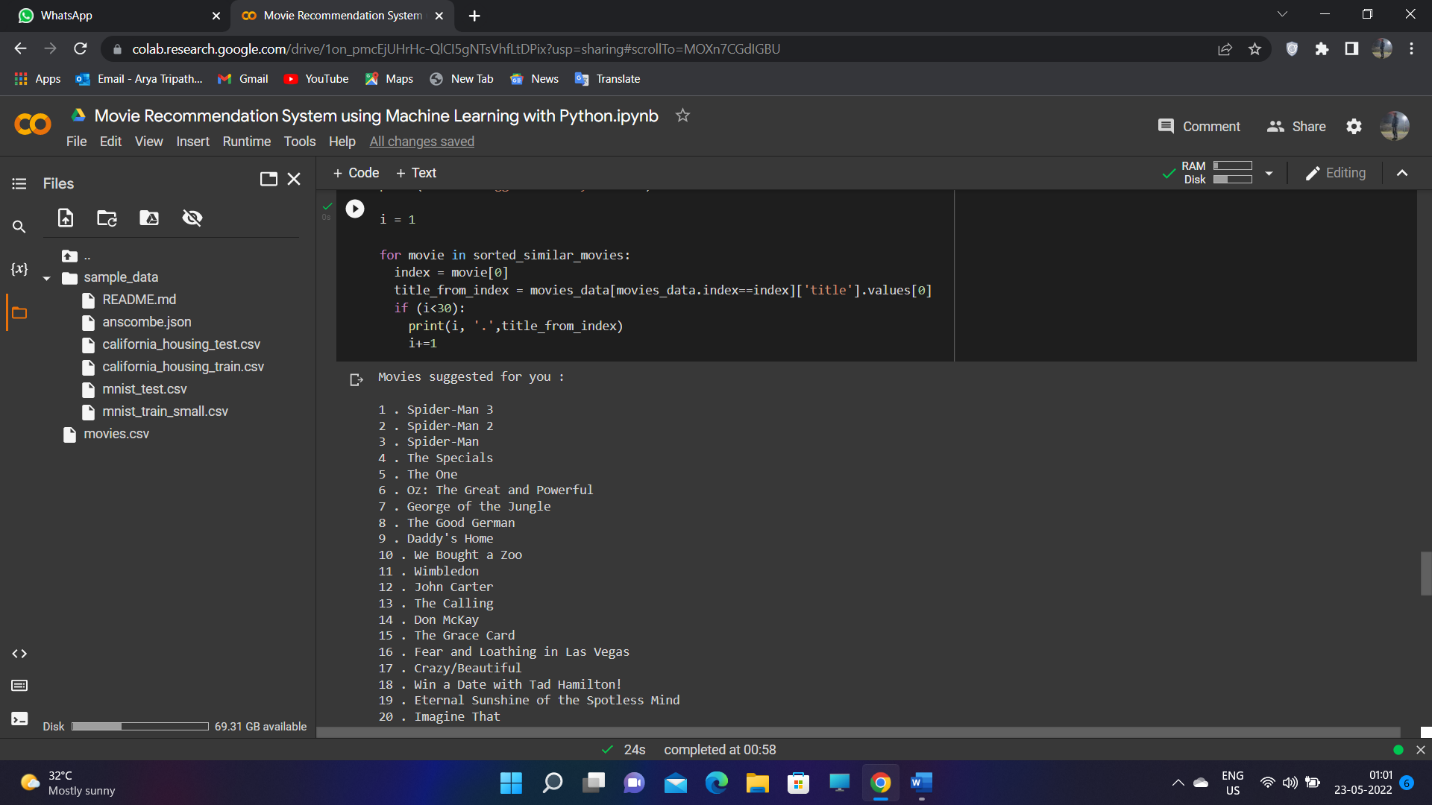
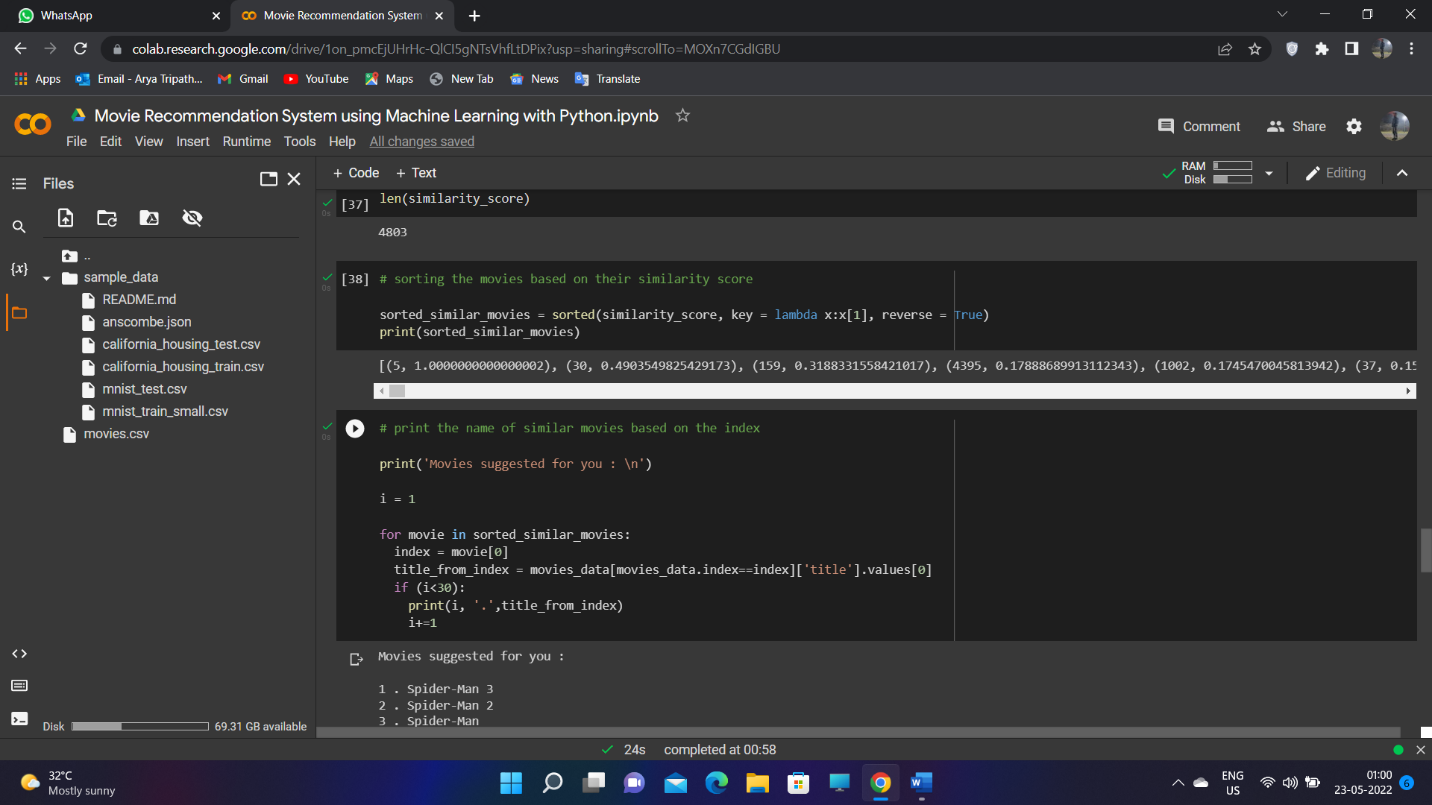
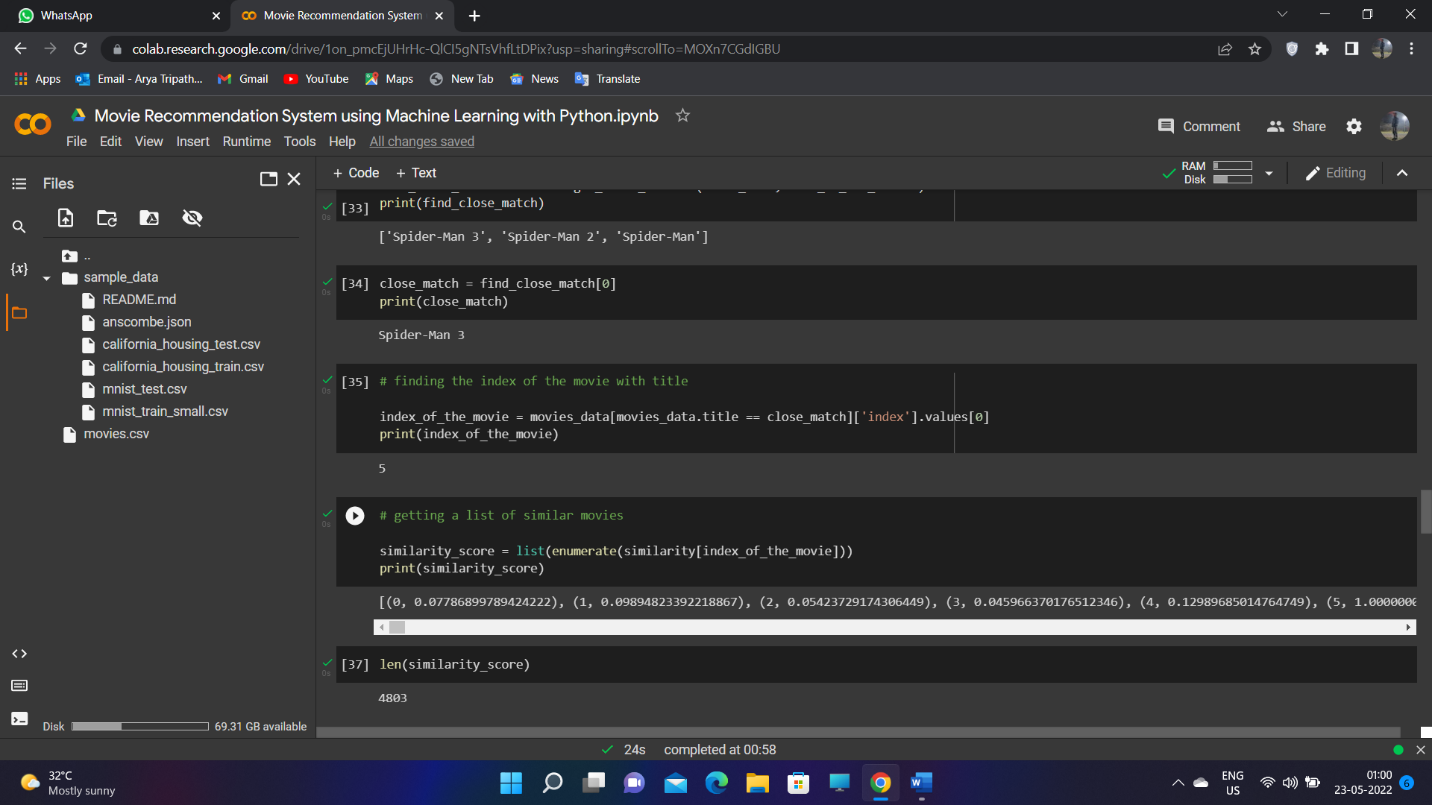
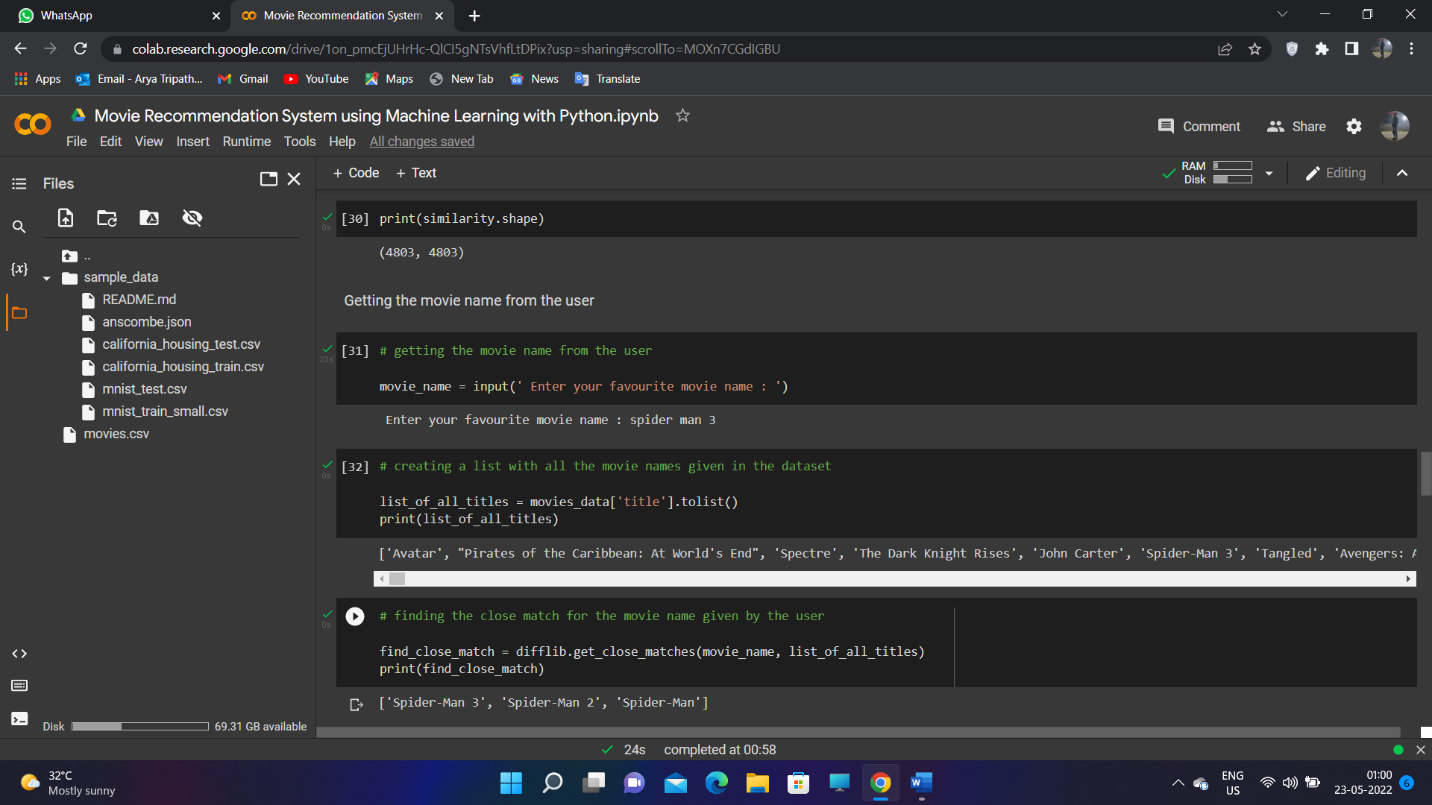
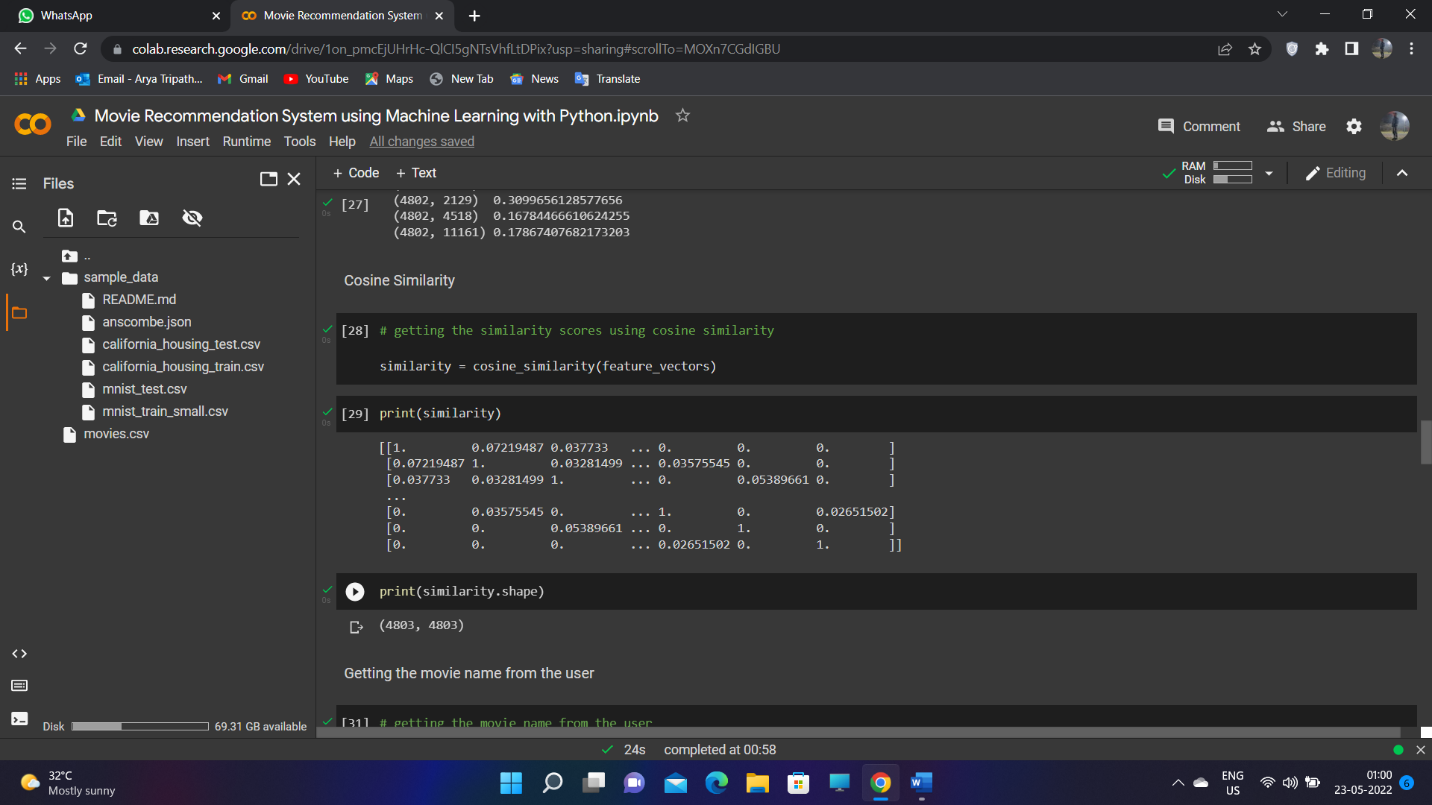
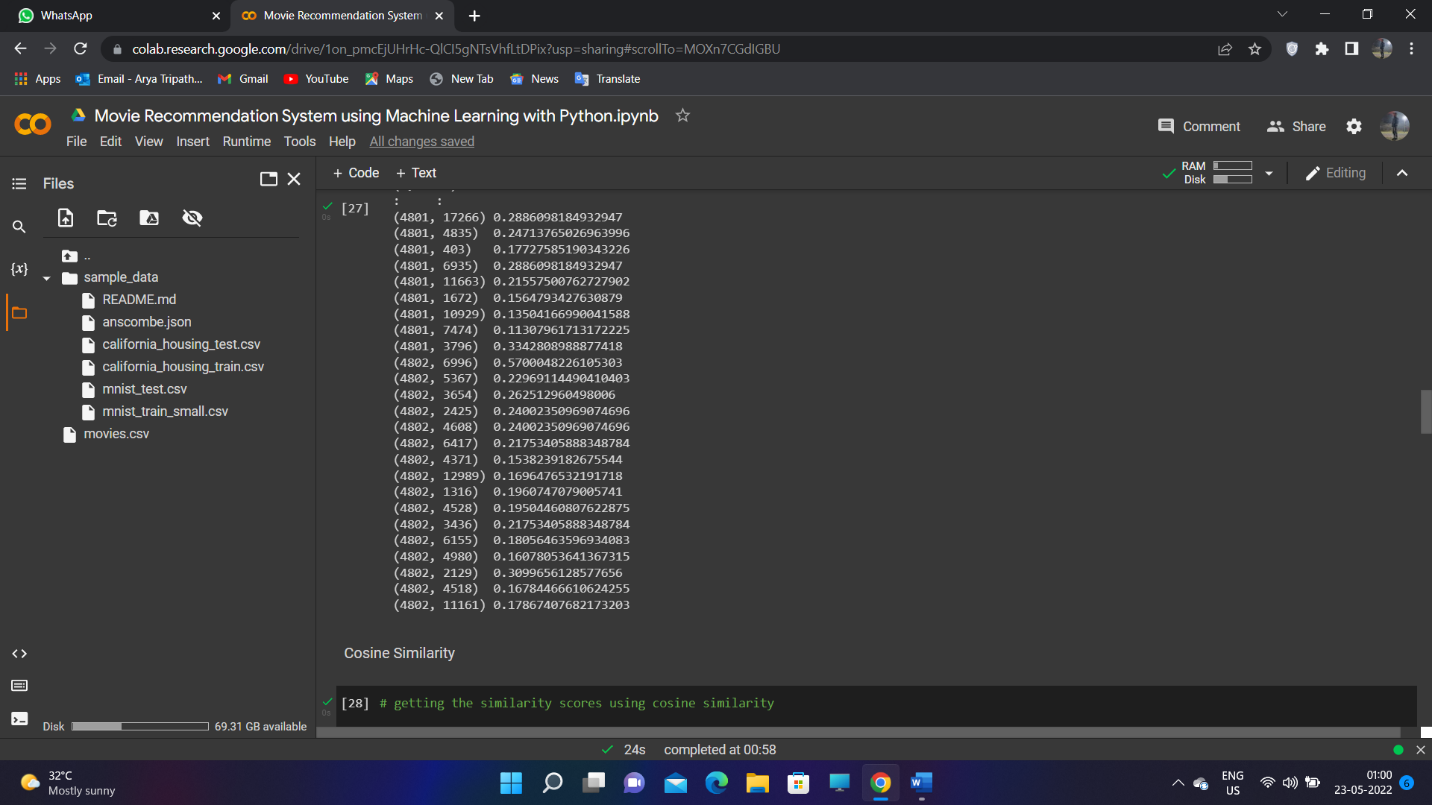
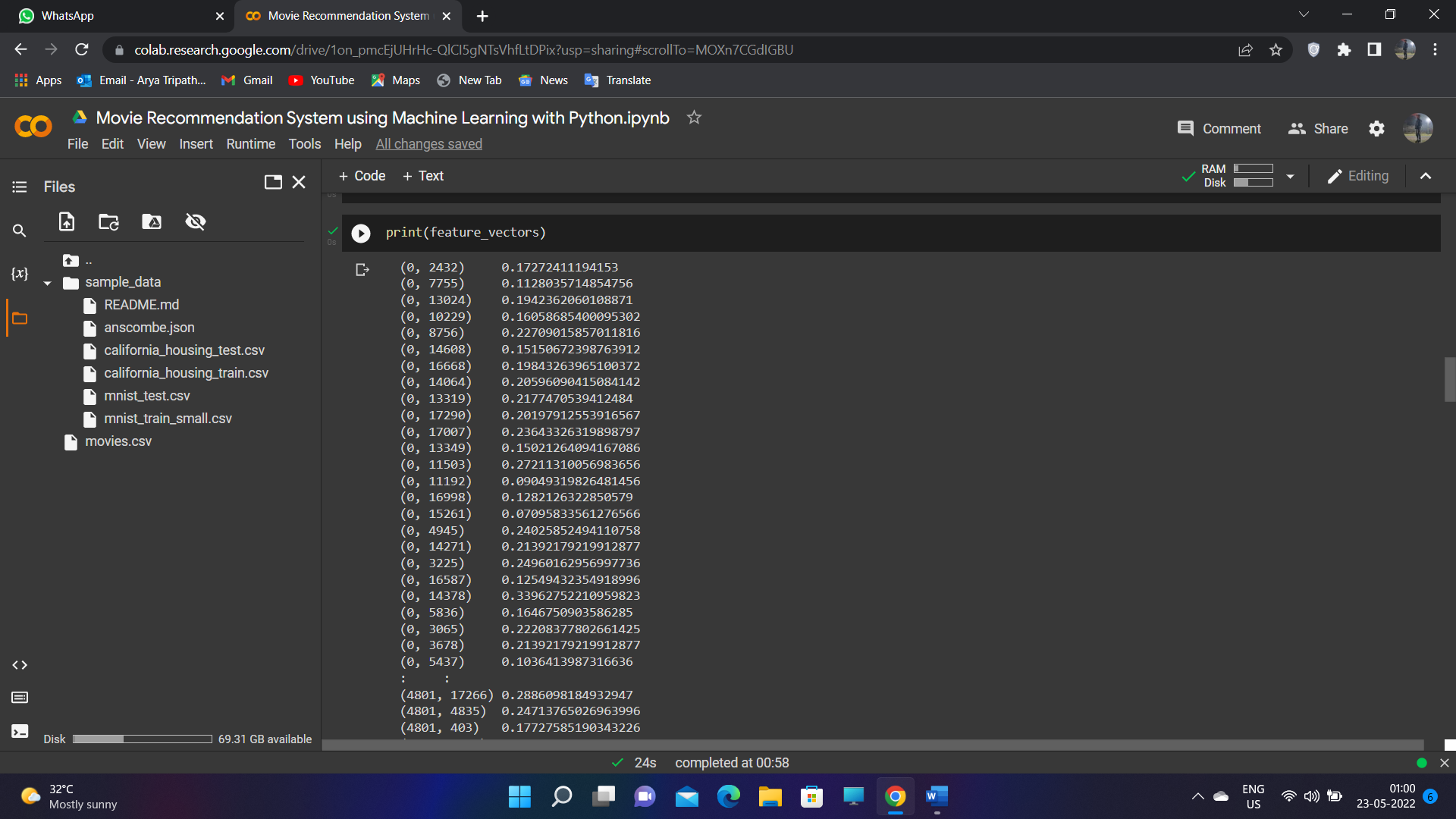
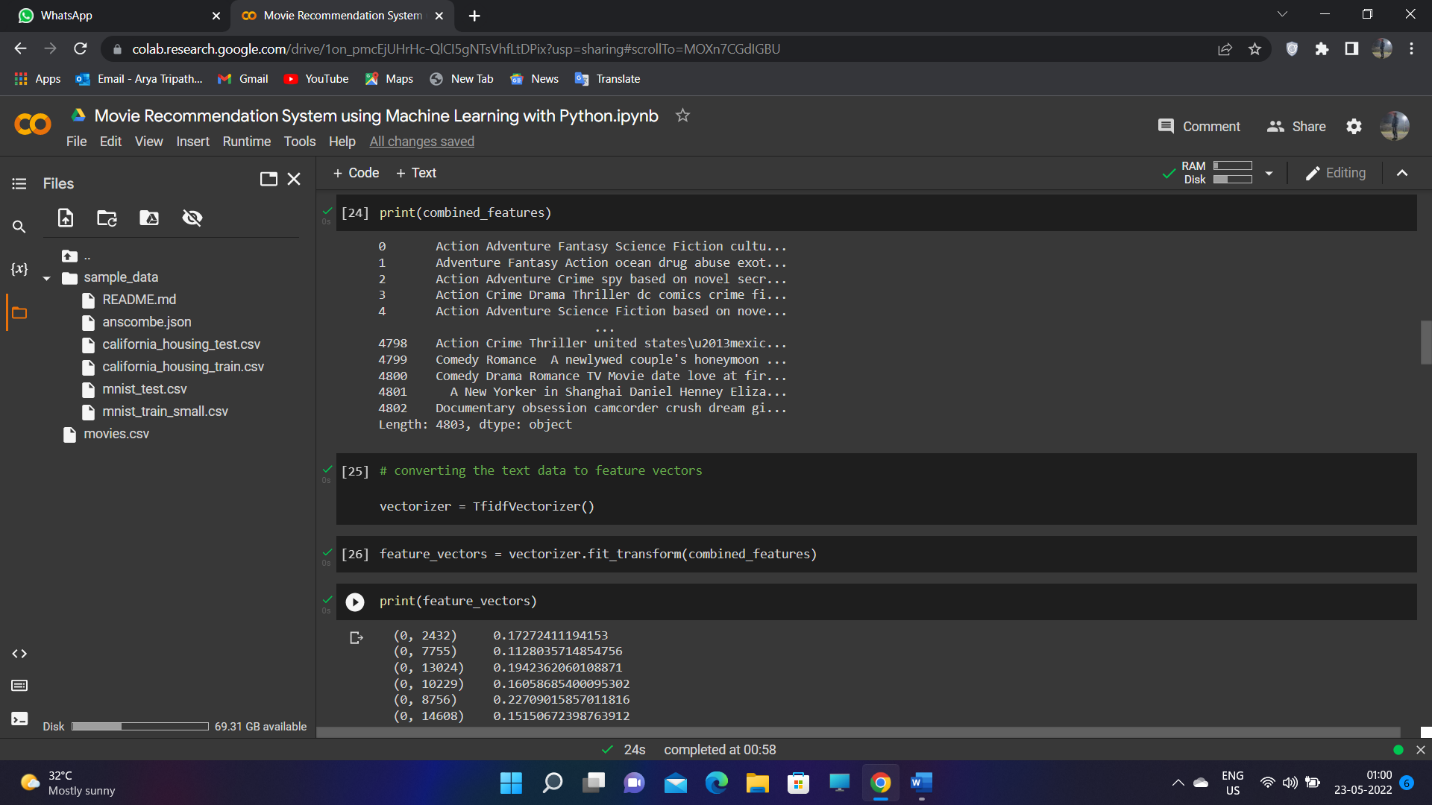
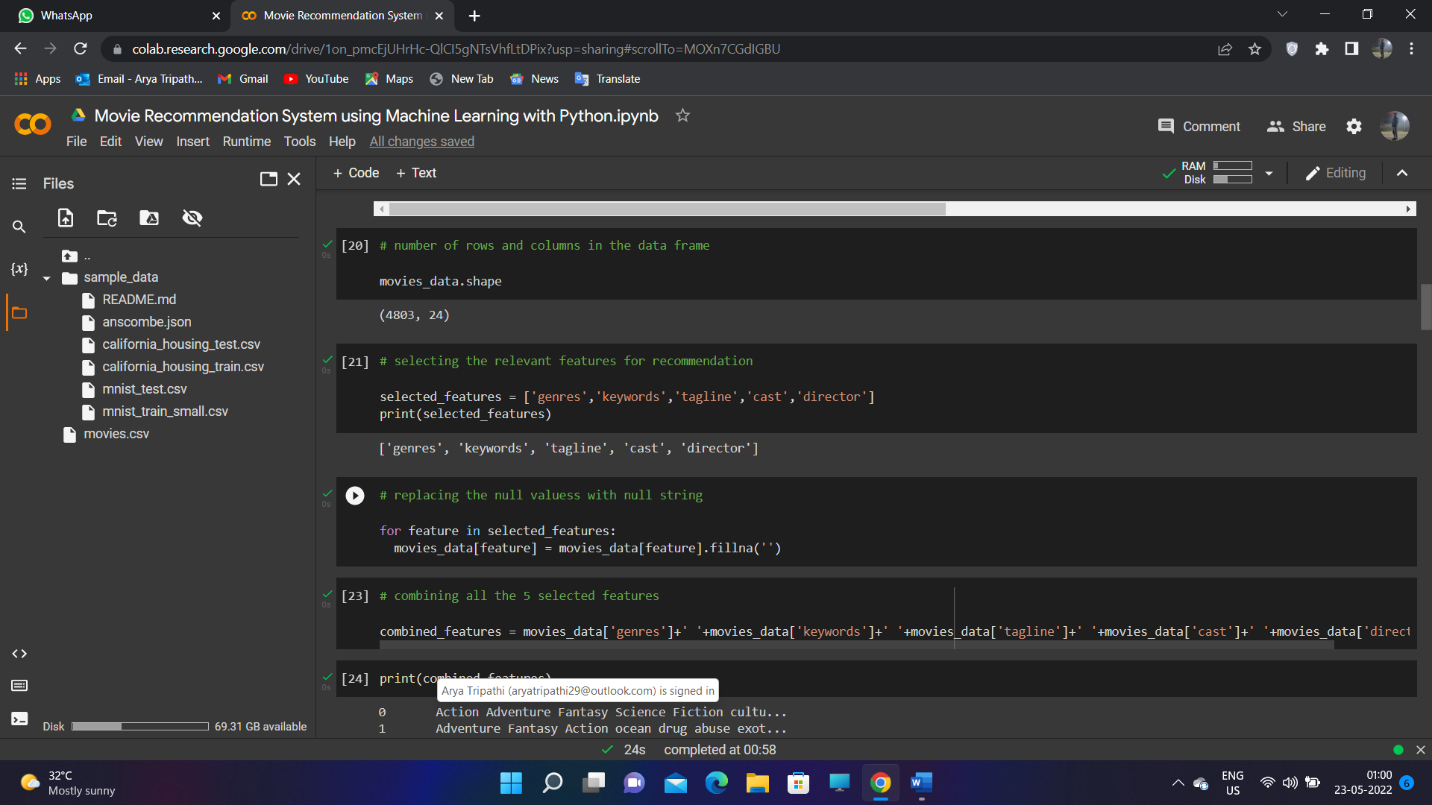
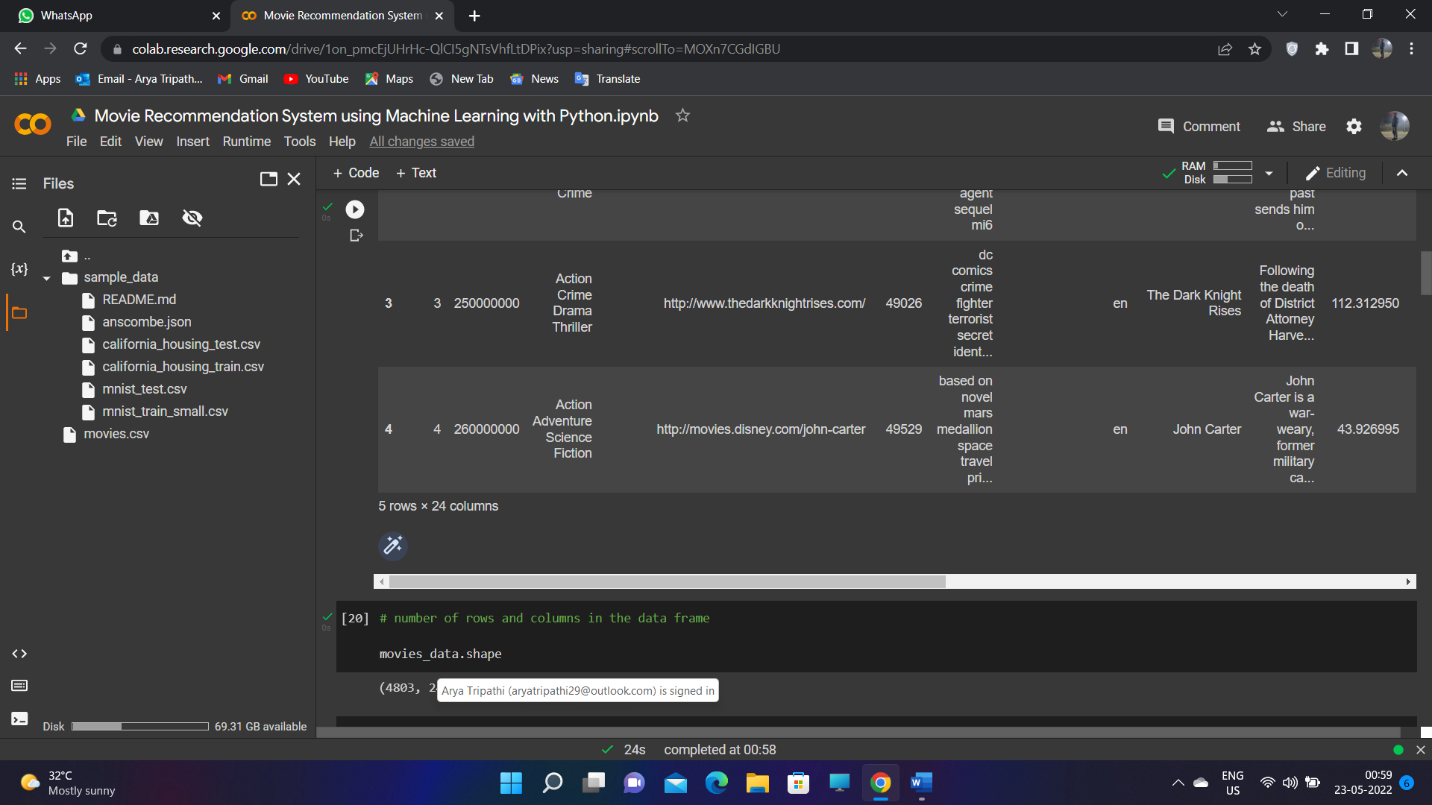
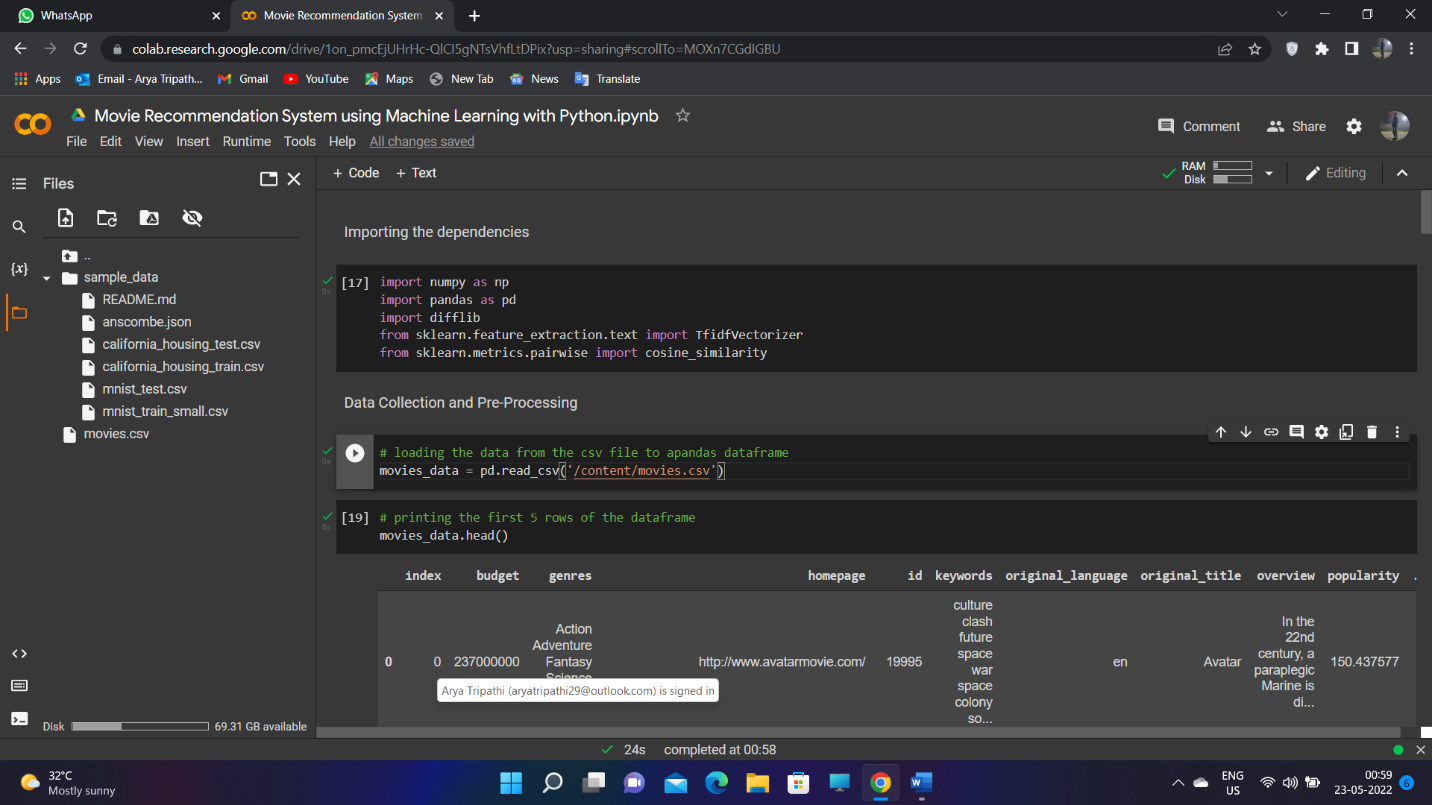
**Analysis of the code: -**

* First, we have imported four libraries – numpy as np, pandas as pd, matplotlib.pyplot as plt and seaborn library as sns.
* Secondly, we have loaded our dataset – tracks.csv using read\_csv() function of pandas library and used head() function for displaying first five rows of the dataset.
* Then, for checking null values in the dataset, we have used isnull() function of pandas library.
* After this, we have created a Convolution Map using heatmap() function of Seaborn library and then , we have created a Regression Plot using regplot() function of Seaborn library.

Then, we used displot() function of Seaborn library, to create a Distribution plot for the ‘number of songs per year’. At last, we used barplot() function of Seaborn library to create a barplot.

****

****

**Screenshots of code: -**

**7**

**8**

**9**

**10**

**11**

**12**

**13**

**Project Outcome: -**

From this project, we learnt to describe a flow process for data science problems and classified data science problems into standard typology. We also learnt about correlating results to the solution approach followed and assessing the solution approach.

**Project Conclusion: -**

In this work, it is been concluded that detection of age and gender take consideration of research few years ago. In this work, technique of morphological and SIFT is applied to search key features from the images. The key features of the images are the color and texture of the image. The simulation results shows that proposed algorithm performed well in terms of fault detection rate and accuracy. In future, further improvement will be done in proposed work for iris reorganization for batter reorganization

**14**