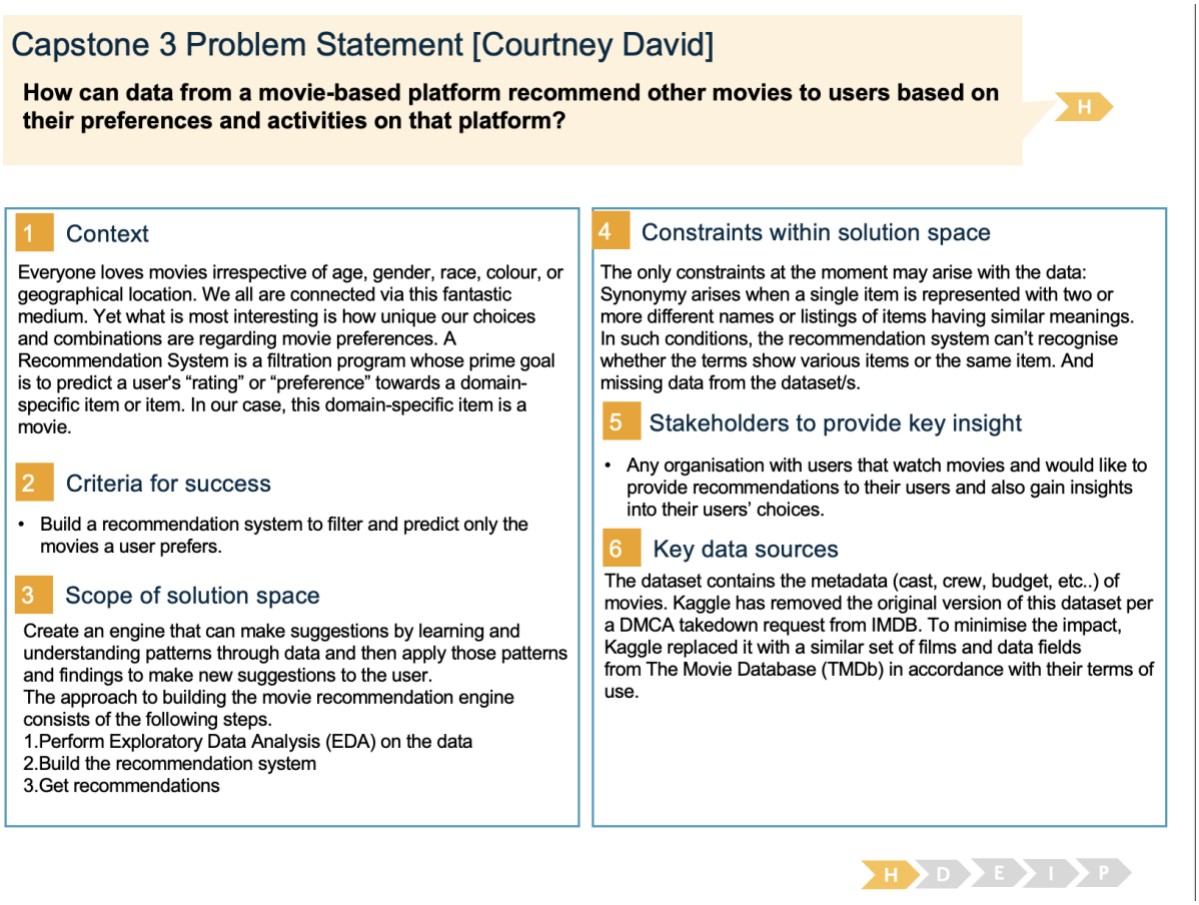
Problem Statement for Capstone 3 Project

Capstone 3 project idea: Build a recommendation system to filter and predict only the movies a user prefers.

Problem Statement: How can data from a movie-based platform recommend other movies to users based on their preferences and activities on that platform?

Problem statement summary



Context

Everyone loves movies irrespective of age, gender, race, colour, or geographical location. We all are connected via this fantastic medium. Yet what is most interesting is how unique our choices and combinations are regarding movie preferences. A Recommendation System is a filtration program whose prime goal is to predict a user's "rating" or "preference" towards a domain-specific item or item. In our case, this domain-specific item is a movie. Recommendation systems also learn your viewing patterns and may suggest content based on the time of the day or your watching patterns. The recommender system helps users quickly find their preferred movies without searching from the extended content catalogue. From a business perspective, the more relevant content, or movies a user finds on any platform, the higher their engagement and, as a result, increased revenue.

Criteria for success

In this project, I will build a recommendation system to filter and predict only the movies a user prefers.
There are three types of recommendation systems.

- Demographic Filtering: The recommendations are the same for every user. They are generalised, not personalised. These types of systems are behind sections like “Top Trending”.
- Content-based Filtering: These suggest recommendations based on the item metadata (movie, product, song, etc.). Here, the main idea is that if a user likes an item, they will also like items like it.
- Collaboration-based Filtering: These systems make recommendations by grouping users with similar interests. For this system, metadata of the item is not required.

In this project, we are building a Content-based recommendation engine for movies.

Dataset

The dataset contains the metadata (cast, crew, budget, etc..) of movies. Kaggle has removed the original version of this dataset per a DMCA takedown request from IMDB. To minimise the impact, Kaggle replaced it with a similar set of films and data fields from The Movie Database (TMDb) in accordance with their terms of use. There will be two datasets used: credits_data and movies_data. Both are CSV files, and a quick overview of datasets is below:

	movie_id	title	cast	crew
0	19995	Avatar	[{"cast_id": 242, "character": "Jake Sully", "...	[{"credit_id": "52fe48009251416c750aca23", "de...
1	285	Pirates of the Caribbean: At World's End	[{"cast_id": 4, "character": "Captain Jack Spa...	[{"credit_id": "52fe4232c3a36847f800b579", "de...
2	206647	Spectre	[{"cast_id": 1, "character": "James Bond", "cr...	[{"credit_id": "54805967c3a36829b5002c41", "de...
3	49026	The Dark Knight Rises	[{"cast_id": 2, "character": "Bruce Wayne / Ba...	[{"credit_id": "52fe4781c3a36847f81398c3", "de...
4	49529	John Carter	[{"cast_id": 5, "character": "John Carter", "c...	[{"credit_id": "52fe479ac3a36847f813eaa3", "de...

	budget	genres	homepage	id	keywords	original_language	original_title	overview	popularity	production_com
0	237000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.avatarmovie.com/	19995	[{"id": 1463, "name": "culture clash"}, {"id":...	en	Avatar	In the 22nd century, a paraplegic Marine is di...	150.437577	[{"name": "In Film Partner
1	300000000	[{"id": 12, "name": "Adventure"}, {"id": 14, "...	http://disney.go.com/disneypictures/pirates/	285	[{"id": 270, "name": "ocean"}, {"id": 726, "na...	en	Pirates of the Caribbean: At World's End	Captain Barbossa, long believed to be dead, ha...	139.082615	[{"name": "Walt Pictures", "id":
2	245000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://www.sonypictures.com/movies/spectre/	206647	[{"id": 470, "name": "spy"}, {"id": 818, "name...	en	Spectre	A cryptic message from Bond's past sends him o...	107.376788	[{"name": "Co Pictures",
3	250000000	[{"id": 28, "name": "Action"}, {"id": 80, "nam...	http://www.thedarkknightrises.com/	49026	[{"id": 849, "name": "dc comics"}, {"id": 853,...	en	The Dark Knight Rises	Following the death of District Attorney Harve...	112.312950	[{"name": "Leg Pictures", "id
4	260000000	[{"id": 28, "name": "Action"}, {"id": 12, "nam...	http://movies.disney.com/john-carter	49529	[{"id": 818, "name": "based on novel"}, {"id":...	en	John Carter	John Carter is a war-weary, former military ca...	43.926995	[{"name": "Walt Pictures",

A potential method to solve the problem

Create an engine that can make suggestions by learning and understanding patterns through data and then apply those patterns and findings to make new suggestions to the user.

The approach to building the movie recommendation engine consists of the following steps.

1. Perform Exploratory Data Analysis (EDA) on the data
2. Build the recommendation system
3. Get recommendations

Deliverables for this project

A GitHub repo containing the work completed for each step of the project, including:

- A slide deck.
- A project report.