Project Activities Summary

|  |  |  |  |
| --- | --- | --- | --- |
| S.No. | Activity Name | Time Spent  (Minutes) | Description |
| 1 | Idea Selection | 240 | This task involved the entire team, consisting of members Diya, Madni Ali and Gaurav. We collectively decided to work on a project involving movie  reviews datasets. |
| 2 | Data Collection | 300 | We aimed to gather three specific datasets: 1. Critics Movie Reviews, 2. Users Movie Reviews, 3. Movies Metadata. I was responsible for the Movies dataset, so that we can merge the user and critic reviews with the movies on movie title, to infer on the same. I have the IMDB movie dataset. It contains movie release data over the years, along with movie title, genres and actors etc. |
| 3 | Setting Up Anaconda Environment | 5 | Used the Anaconda environment provided by other members to keep the environment the same. |
| 5 | Analyzing the Datasets | 30 | The IMDB movie dataset contains 900,000 rows and 25 columns. This has a lot of columns many were irrelevant for our analysis and a few were of work, that I cleaned. |
| 6 | Dropping Columns from Datasets | 5 | Given the large volume of data, I opted to drop several columns before importing the datasets into our database to manage space more efficiently. |
| 7 | Analysis and Visualization | 150 | Conducted detailed analyses and generated various visualizations to represent the sentiment trends of the  reviews over time. |
| 8 | Combining Datasets | 90 | This step involved merging all three |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  | datasets into a single IPython notebook (ipynb). This was a challenging task as we needed to merge all the datasets. It was a collaborative effort by all team members. |
| 9 | Saving in Database | 40 | This was a time consuming and intricate task, since I needed to check later is the sentiments of each movie have been saved accurately and that each movie has as many reviews both from critic and user that are enough for the analysis. |
| 10 | Findings | 50 | From that, I combined and preprocessed all the data from both datasets and then analyzed it. First, I made visualization graph on the combination of the three datasets, about the influence and trends of review analysis. One improvement was normalizing the data by percentage and not by counts. Created visualizations to display the insights derived from our analyses of the combined dataset. |
| 11 | Drawing the Final Result | 30 | This was the main objective of the project, so It was the gist and quiet an important task. I inferred from revies, how they have affected the rise or fall of the movie release, and I did manage to find a good impact of positive reviews on the recent movie releases, covering the requirement of our research. |
| 12 | Research Paper | 150 | I have worked on the methodology of the report, elaborating the working and utilization of the model we have entailed. And I worked on the Evaluation, since I was responsible for the inference of the Influence and Impacts of the reviews, i.e. the final data analysis task. |