**Phase 2 Process/Contribution Memo**

Name: Akash Gajendra

UIN: 927009622

**Part A**

Last week, we prepared our design document which conceptualized a database using 5 separate entities - Analyst, User, Movie, Actor and Director - each of which had their own set of relationships and fields.

Upon receiving the data, we kickstarted our discovery phase where we tried to understand the datasets on a deeper level by analyzing the kinds of fields and entries they contained. After this analysis, it became clear that our assumptions about the data during the design document phase had been inaccurate to a certain degree we reconstructed our entity relationship diagram to reflect the datasets we obtained. The changes have been listed design changes section of the document. Upon completion of the discovery phase, we began the data-preprocessing phase where we cleaned and formatted the data according to our database needs. These are some **but not all** the changes that we performed to the files

1. crew.csv – Delete first column, Check for missing movie entry
2. names.csv – Delete first column, Check for duplicates
3. principals.csv – Delete first column, Check for duplicate entries with same Title ID & Name ID & Category and delete the duplicates
4. customer\_rating.csv – Delete first column, Check for duplicate reviews by the user for the same movie and delete the earlier one
5. titles.csv – Delete first column, Ignore ‘year’, Check for duplicates

In general, our approach to pre-processing consisted primarily of getting rid of duplicates, find missing values and looking for null values. After this phase was completed, we created the database and populated all the entities. Along with this, we sketched out the final design for application GUI.

**Changes:** Following the inspection of the datasets to be used, we were forced to revise our entity relationship diagram as our base assumptions from the previous case were not met with the new data.

* The Analyst entity previously has an attribute with a list of “User IDs” which has now been replaced by a list of IDs from the customer\_ratings.csv file which now makes up a new customer entity.
* The Movie entity was replaced by the Titles entity which draws from the titles.csv file. Appropriate data types were assigned and used for the attributes.
* The major differences arise in the remaining entities – In the previous case, we had an actor & director entity that tied back to the Movies entity using their own unique IDs. However, since we received a names.csv file with all the name IDs for the individuals that make up the crew for the movies, we had to create a new entity that stored the personal information associated with the name IDs being used in other entities.
* The names.csv was used to construct the names entity and this was connected using Name Id to two other entities named – Crew and Principals – Both of these consist of movie staffing information drawn from crew.csv and principals.csv respectively.
* All the relationship information can be found on Figure 2 near the top right corner.

Please find below the diagrams showcasing both ERDs.

**Diagram

Description automatically generated**

**Figure 1: Old Entity Diagram**

**Diagram, schematic

Description automatically generated**

**Figure 2: Updated Entity Diagram**

**Part B**

Our process for designing the GUI consisted primarily of three phases. The first phase consisted of listing out potential features for both the user/analyst style and sketching a skeleton layout for the GUI, both of which we performed in the class activity last week.

The next phase, our team constructed our wireframe designs using Figma (UI/UX visualization tool) and finalized the color scheme for the different frames.

The last phase consisted of adding error handling capabilities to the UI. For example, upon entering wrong credentials the UI will display incorrect password/username entered in either of the User/Analyst login portal. That’s just one of many error handling methods we have incorporated into our GUI design.

**Part C**

I was initially involved with phase 1 design of the UI for the analyst and viewer modes i.e building the base layout for each of the wireframes we were constructing. After the prototype design was submitted, I was primarily involved with the database side of the project. I analyzed the structure of the datasets and updated our ERD to reflect the changing relationships. Moreover, I was involved with the data-preprocessing of the many of the csv files as they required formatting/cleaning for successful population. Along with this, I helped with making and testing the initial versions of the population script as well.