In this phase, you will create the database, database tables, and the corresponding relationships in MySQL. You will populate the tables in MySQL. Finally, you will write at least 8 business questions which can be answered by writing SQL queries. You do not have to write these queries in this phase, but your questions should be such that they can be answered using SQL queries. Also, your questions and queries should be of varying complexities.

Your deliverable will be a three to four-page report containing the following information:

1. Discussion of how you converted the dataset into tables.

2. Challenges faced during importing of your data and how did you overcome these data importation challenges.

3. A complete data dictionary for every table in your database.

4. The list of business questions.

1. **Discussion of how you converted the dataset into tables.**
2. **Initial Data Source:**
   * The dataset was sourced from [Kaggle](https://www.kaggle.com/datasets/gulsahdemiryurek/harry-potter-dataset), it contains the information from the first 3 Harry Potter movies.
3. **Creation of Tables as per ERD:**
   * The raw data was organized into several distinct tables, each focusing on different elements from the Harry Potter universe.

**a. Potions Table:**

* + This table cataloged information about potions, including their names, ingredients, effects, and characteristics.

**b. Hogwarts Staff Table:**

* + It contained details about the Hogwarts staff, including job titles, loyalty, wands, blood status, and house affiliations.

**c. Patronus Table:**

* + This table listed various Patronus, providing details like the name of each Patronus.

**d. Wizard Table:**

* + It included comprehensive information about wizards and witches, such as names, jobs, physical attributes, and life spans.

**e. Spells Table:**

* + This table captured data on spells, including their names, incantations, types, effects, and associated light.

**f. Movie Dialogues Table:**

* + It focused on dialogues from the movies, including the names of characters and their lines.

**g. Hogwarts Student Table:**

* + Details about Hogwarts students were listed here, including blood status, loyalty, and wands.

1. **Conversion to CSV Format:**
   * Each table was converted into CSV format for ease of import into MySQL.
2. **Importing into SQL:**
   * The CSV files were imported into MySQL, which involved structuring the data according to the database schema.
3. **Data Inspection and Key Assignments:**
   * After import, each table was checked for accuracy and consistency. Primary keys were assigned for unique identification, and foreign keys were used to link related tables. The unique column ‘Id’ was chosen as the primary key for the wizard table and was referenced as foreign key for the others.
4. **Challenges faced during importing of your data and how did you overcome these data importation challenges.**

When importing the Harry Potter dataset into MySQL, I faced a couple of key challenges, which I managed to successfully navigate.

Firstly, I had to deal with the issue of redundant data in the Excel files. To tackle this, I thoroughly reviewed the dataset before importing it. This step was vital to remove any duplicate or unnecessary data, making the dataset more efficient for storage and querying in the SQL database. Additionally, ensuring the correct character encoding was a critical step. This was important to avoid any unusual characters appearing in the database, which could affect the clarity and accuracy of the data.

Another challenge was related to the format of data in the CSV files. I noticed that some columns had extra quotation marks and special characters, which could cause problems during the import process. To solve this, I adjusted the character encoding settings to match MySQL's import requirements. This approach effectively fixed the formatting issues, leading to a smooth and error-free import of the data into MySQL.

Snippet of the database schema:

A screenshot of a computer

Description automatically generated

1. **Data Dictionary**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table** | **Field** | **Type** | **Description** | **Data Type** |
| **Potions** | Name | PK | Unique Identifier for each spell | Text |
|  | Known Ingredients |  | The ingredients known to be used in the potion. | Text |
|  | Effect |  | The result or impact of using the potion. | Text |
|  | Characteristics |  | Distinctive features or properties of the potion. | Text |
|  | Difficulty Level |  | The level of difficulty in brewing the potion. | Text |
| **Hogwarts Staff** | Id | PK/FK | Unique identifier for each staff member. | Double |
|  | Job Title |  | The title or position held by the staff member at Hogwarts. | Text |
|  | Loyalty |  | The entity or person to whom the staff member is loyal. | Text |
|  | Wand |  | Description of the staff member's wand. | Text |
|  | Blood Status |  | The blood status (pureblood, half-blood, muggle-born) of the staff member. | Text |
|  | House |  | The Hogwarts house with which the staff member is associated. | Text |
| **Patronus** | Id | PK/FK | Unique identifier for each Patronus charm. | Double |
|  | Patronus Name |  | The name or type of the Patronus. | Text |
| **Wizard** | Id | PK | Unique identifier for each individual. | Double |
|  | Full Name | CK | The complete name of the individual. | Text |
|  | Fname |  | First name of the individual. | Text |
|  | Middle Name |  | Middle name of the individual. | Text |
|  | Lname |  | Last name of the individual. | Text |
|  | Gender |  | Gender of the individual. | Text |
|  | Job |  | The individual's occupation. | Text |
|  | Hair Color |  | Hair color of the individual. | Text |
|  | Eye Color |  | Eye color of the individual. | Text |
|  | Birth |  | Birth date of the individual. | Text |
|  | Death |  | Death date of the individual. | Text |
|  | Species |  | The species (e.g., human, Werewolf, Half-Human) of the individual. | Text |
| **Spells** | Name | PK | The name of the spell. | Text |
|  | Incantation |  | Words needed to conjure the spell | Text |
|  | Type |  | The category or type of the spell. | Text |
|  | Effect |  | What the spell does | Text |
|  | Light |  | Light the spell casts | Text |
| **Movie Dialogues** | Fname |  | First name of the character who delivers the dialogue. | Text |
|  | Lname |  | Last name of the character who delivers the dialogue. | Text |
|  | Sentence |  | The dialogue or sentence spoken in the movie. | Text |
| **Hogwarts Student** | ID | PK/FK | Unique identifier for each student. | Double |
|  | Blood Status |  | The blood status (pureblood, half-blood, muggle-born) of the student. | Text |
|  | Loyalty |  | The entity, person, or cause to which the student is loyal. | Text |
|  | Wand |  | Description of the student's wand. | Text |

1. **Business Questions:**
2. **Family Lineage and House Sorting:**

Do members of wizarding families sharing the same surname tend to be sorted into the same Hogwarts house and share loyalty alignments, suggesting a lineage-based influence on house sorting and affiliations?

1. **Character and House Alignment:**

Determine the distribution of characters, including staff and students, among the four Hogwarts houses and identify any significant attributes (like Loyalty or Blood Type) that may influence house assignment?

1. **Potion Complexity Analysis and Usage:**

How does the number of known ingredients in a potion relate to its described effect and characteristics? Can we identify a complexity rating for each potion? Then, determine if more complex potions are mentioned more or less frequently across the Harry Potter series files.

1. **House Diversity Analysis:**

What is the distribution of blood status (pureblood, half-blood, muggle-born) among students in each Hogwarts house, based on data from Hogwarts Student.csv and House Name.csv?

1. **Movie Dialogues and Character Representation:**

How many unique dialogues are delivered by characters in the movies? Can we analyze the representation of blood status and loyalty in the dialogues spoken by characters?

1. **Staff Loyalty and Characteristics:**

How does the blood status of Hogwarts staff members correlate with their job titles, and what are the distinct characteristics of their wands?

1. **Life Events and Blood Status:**

How do birth and death rates vary among wizards based on their blood status, and are there any notable patterns or trends?

1. **Popular Characters:**

Among the characters featured in all three movies, which are the top 3 that have garnered the highest level of popularity and major screen time, as measured by their frequency of dialogue throughout the 3 movies?