BAX 421 Report

Phase 3: Database creation, table population, and business questions

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

To convert the datasets into database tables, I first examined the CSV files' formats. Among the seven CSV files, each represents data from distinct regions: Global, Korea, North America, Europe, Taiwan, Japan, and an aggregate of all regions at the diamond rank. These files shared a same columnar framework, featuring fields like Champion, Games Played, KDA, Win Rate, Pick Rate, Ban Rate, CS, and Gold. This suggested that a standardized table design could be applied across the various sets of regional data. I decided to focus on the data from Korea, North America, Europe, Taiwan, and Japan to create a more effective comparative analysis across these regions. By implementing database normalization principles, I made a streamlined schema that minimized data duplication and preserved the integrity of the data. This process included the creation of a Regions table to catalog the unique identifiers and corresponding names of each region, along with a Champions table that consolidated shared champion data. For region-specific statistics, I devised individual ChampionStats tables for EUW, NA, JP, TW, and KR, making sure that each of these tables linked back to the Champions table through a foreign key. So this will allow for the orderly conversion of the raw CSV data into a coherent set of relational tables within the MySQL environment.

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

1.The sheer volume and complexity of the data posed the first major challenge. Multiple CSV files from different regions containing thousands of rows made the import process daunting.

Solution:

To manage this, I used Excel to carefully examined each csv file, And I manually fixed the column to make it fit better to mysql. This not only made the import process more efficient, but also reduced the risk of overload.

2. Associating championship statistics with the correct Champion ID and Region ID was tricky due to the limitations of foreign keys. Initially, this led to many import errors.

Solution:

To resolve this issue, I first imported the Champion and Region data, assigning unique IDs to each.I then updated the Champion statistics to include these IDs. This involved mapping the Champion names and Regions to their respective IDs using a lookup strategy in the script.

As a newcomer to MySQL, I was not familiar with SQL syntax and it took me much more time to figure out how to process it. In addition, it was also complex to ensure data integrity and maintaining the relational structure of the database. To overcome these challenges, I utilized batch processing and scripting to automate and simplify the import process. I also took the time to learn MySQL commands and understand the complexities of database administration.

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

Regions Table

|  |  |  |  |
| --- | --- | --- | --- |
| Column Name | Data Type | Contraints | Description |
| Region ID | INT | PRIMARY KEY, AUTO INCREMENT | The ID for each region |
| RegionName | VARCHAR(255) | NOT NULL, UNIQUE | The name of the region |

Champions Table

|  |  |  |  |
| --- | --- | --- | --- |
| Column Name | Data Type | Contraints | Description |
| ChampionID | INT | PRIMARY KEY, AUTO INCREMENT | The ID for each champion |
| ChampionName | VARCHAR(55) | NOT NULL, UNIQUE | The name of the champion |

ChamptionStats\_EUW Table

|  |  |  |  |
| --- | --- | --- | --- |
| Column Name | Data Type | Constraints | Description |
| ChampionStatsID | INT | PRIMARY KEY, AUTO INCREMENT | The identifier for each entry |
| ChampionID | INT | FOREIGN KEY, NOT NULL | Champion ID from the Champions Table |
| GamesPlayed | INT | NOT NULL | Number of Games being played |
| KDA | DECIMAL(10,2) | NOT NULL | The ratio of Kill- Death- Assist |
| WinRate | DECIMAL(5,5) | NOT NULL | The percentage of winning the games |
| PickRate | DECIMAL(5,5) | NOT NULL | The percentage of getting picked |
| BanRate | DECIMAL(5,5) | NOT NULL | The percentage of getting banned |
| CS | DECIMAL(10,2) | NOT NULL | The average number of minion killed |
| Gold | DECIMAL(10,2) | NOT NULL | The average gold earned |
| RegionID | INT | FOREIGN KEY, NOT NULL | RegionID from the Regions table |

ChamptionStats\_NA Table

|  |  |  |  |
| --- | --- | --- | --- |
| Column Name | Data Type | Constraints | Description |
| ChampionStatsID | INT | PRIMARY KEY, AUTO INCREMENT | The identifier for each entry |
| ChampionID | INT | FOREIGN KEY, NOT NULL | Champion ID from the Champions Table |
| GamesPlayed | INT | NOT NULL | Number of Games being played |
| KDA | DECIMAL(10,2) | NOT NULL | The ratio of Kill- Death- Assist |
| WinRate | DECIMAL(5,5) | NOT NULL | The percentage of winning the games |
| PickRate | DECIMAL(5,5) | NOT NULL | The percentage of getting picked |
| BanRate | DECIMAL(5,5) | NOT NULL | The percentage of getting banned |
| CS | DECIMAL(10,2) | NOT NULL | The average number of minion killed |
| Gold | DECIMAL(10,2) | NOT NULL | The average gold earned |
| RegionID | INT | FOREIGN KEY, NOT NULL | RegionID from the Regions table |

ChamptionStats\_KR Table

|  |  |  |  |
| --- | --- | --- | --- |
| Column Name | Data Type | Constraints | Description |
| ChampionStatsID | INT | PRIMARY KEY, AUTO INCREMENT | The identifier for each entry |
| ChampionID | INT | FOREIGN KEY, NOT NULL | Champion ID from the Champions Table |
| GamesPlayed | INT | NOT NULL | Number of Games being played |
| KDA | DECIMAL(10,2) | NOT NULL | The ratio of Kill- Death- Assist |
| WinRate | DECIMAL(5,5) | NOT NULL | The percentage of winning the games |
| PickRate | DECIMAL(5,5) | NOT NULL | The percentage of getting picked |
| BanRate | DECIMAL(5,5) | NOT NULL | The percentage of getting banned |
| CS | DECIMAL(10,2) | NOT NULL | The average number of minion killed |
| Gold | DECIMAL(10,2) | NOT NULL | The average gold earned |
| RegionID | INT | FOREIGN KEY, NOT NULL | RegionID from the Regions table |

ChamptionStats\_JP Table

|  |  |  |  |
| --- | --- | --- | --- |
| Column Name | Data Type | Constraints | Description |
| ChampionStatsID | INT | PRIMARY KEY, AUTO INCREMENT | The identifier for each entry |
| ChampionID | INT | FOREIGN KEY, NOT NULL | Champion ID from the Champions Table |
| GamesPlayed | INT | NOT NULL | Number of Games being played |
| KDA | DECIMAL(10,2) | NOT NULL | The ratio of Kill- Death- Assist |
| WinRate | DECIMAL(5,5) | NOT NULL | The percentage of winning the games |
| PickRate | DECIMAL(5,5) | NOT NULL | The percentage of getting picked |
| BanRate | DECIMAL(5,5) | NOT NULL | The percentage of getting banned |
| CS | DECIMAL(10,2) | NOT NULL | The average number of minion killed |
| Gold | DECIMAL(10,2) | NOT NULL | The average gold earned |
| RegionID | INT | FOREIGN KEY, NOT NULL | RegionID from the Regions table |

ChamptionStats\_TW Table

|  |  |  |  |
| --- | --- | --- | --- |
| Column Name | Data Type | Constraints | Description |
| ChampionStatsID | INT | PRIMARY KEY, AUTO INCREMENT | The identifier for each entry |
| ChampionID | INT | FOREIGN KEY, NOT NULL | Champion ID from the Champions Table |
| GamesPlayed | INT | NOT NULL | Number of Games being played |
| KDA | DECIMAL(10,2) | NOT NULL | The ratio of Kill- Death- Assist |
| WinRate | DECIMAL(5,5) | NOT NULL | The percentage of winning the games |
| PickRate | DECIMAL(5,5) | NOT NULL | The percentage of getting picked |
| BanRate | DECIMAL(5,5) | NOT NULL | The percentage of getting banned |
| CS | DECIMAL(10,2) | NOT NULL | The average number of minion killed |
| Gold | DECIMAL(10,2) | NOT NULL | The average gold earned |
| RegionID | INT | FOREIGN KEY, NOT NULL | RegionID from the Regions table |

The list of business questions.

1. Which region has the most balanced gameplay in terms of average win rate, KDA, pick rate, and ban rate across all champions?**Compare average KDA, WinRate, and PickRate for each champion across all regions.**
2. **How do champions rank in terms of overall performance when considering average KDA, win rate, and pick rate across all regions?**
3. **Which champion is most favored across all regions, and how does this preference correlate with their win rate and KDA?**
4. **How does the average gold earned by champions in the TW region compare across different win rate brackets?**
5. **In the EUW region, which champions most frequently achieve a KDA above 5, and how does this reflect on their overall performance?**
6. **Which champion in the KR region has the highest average CS**
7. **List the difference in the average KDA between champions in the EUW and NA regions?**
8. How does the pick rate of champions in the TW region correlate with their games played, and which champions are picked frequently but rarely banned?