# Project Report

**Contributors** : **Eric Kleppen and Bimi Bisht.**

The purpose of this project was to create a relational database in MySQL to store data about Marvel and DC comic book characters and their superpowers. We transformed the data and loaded it using Python. We also checked the relation between eye color and good/bad characters.

**Extract:**

Original Data Sources include:

1. FiveThirtyEight Comic Characters Dataset from Kaggle.com

2. API pull from comicvine.gamespot.com/api/

The superpowers were extracted from the API using a for loop nested within a for loop due to the way the JSON data was structured. Because of API limits, we pulled only the first 100 superpowers from the API.

**Transform**

The data was reviewed, and the irrelevant columns were dropped. The clean data was then used to create data frames that matched the relational schema we designed. We created separate dataframes for the characters and character attributes to maintain a relationship. The real names and the character names were split into separate columns so the characters could later be matched to the superpowers dataset.

**Load**

Each dataframe was loaded into its corresponding table in our schema. We loaded the following datasets:

* Characters
* Character attributes
* Powers
* Character Powers

Once the data was loaded, tests were run to validate that the data was loaded and available for data analysis. Database comic\_db was created, with the following tables:

**CREATE DATABASE comic\_db;**

**USE comic\_db;**

**CREATE TABLE characters (**

**character\_id INT PRIMARY KEY**

**,character\_name VARCHAR(100)**

**,universe VARCHAR(100)**

**);**

**CREATE TABLE char\_attributes (**

**attribute\_id INT PRIMARY KEY**

**,real\_name NVARCHAR(100)**

**,character\_id INT**

**,universe NVARCHAR(25)**

**,ALIGN NVARCHAR(25)**

**,EYE NVARCHAR(25)**

**,SEX NVARCHAR(25)**

**,Year INT**

**,FOREIGN KEY (character\_id) REFERENCES characters(character\_id)**

**);**

**CREATE TABLE powers (**

**power\_id INT PRIMARY KEY**

**,power NVARCHAR(100)**

**);**

**CREATE TABLE char\_powers (**

**power\_id INT**

**,character\_id INT**

**,FOREIGN KEY (power\_id) REFERENCES powers(power\_id)**

**,FOREIGN KEY (character\_id) REFERENCES characters(character\_id)**

**)**

**CREATE VIEW v\_comic\_char\_data**

**AS**

**SELECT c.universe**

**,c.character\_id**

**,c.character\_name**

**,a.attribute\_id**

**,a.real\_name**

**,a.ALIGN**

**,a.EYE**

**,a.SEX**

**,a.Year**

**,p.power\_id**

**,p.power**

**FROM characters c**

**LEFT JOIN char\_attributes a ON a.character\_id = c.character\_id**

**LEFT JOIN char\_powers cp ON cp.character\_id = c.character\_id**

**LEFT JOIN powers p ON p.power\_id = cp.power\_id;**

**-- year in which most characters appeared**

**CREATE VIEW most\_appearances**

**AS**

**SELECT Year**

**,count(\*)**

**FROM char\_attributes**

**WHERE Year IS NOT NULL**

**GROUP BY Year**

**ORDER BY count(\*) DESC;**

**SELECT \***

**FROM most\_appearances;**

**-- Which eye color male or female had more good or bad characters.**

**CREATE VIEW eye\_color**

**AS**

**SELECT a.ALIGN**

**,a.SEX**

**,a.EYE**

**,count(\*) AS "Count"**

**FROM char\_attributes a**

**WHERE a.SEX IS NOT NULL**

**AND a.ALIGN IS NOT NULL**

**AND a.EYE != 'Auburn Hair'**

**GROUP BY a.ALIGN**

**,a.SEX**

**-- , a.EYE**

**ORDER BY a.ALIGN**

**,a.SEX;**

**SELECT \***

**FROM eye\_color;**

**Analysis:** We did not have information about the superpowers of all the comic characters so we chose this particular topic. MySQL database is easy to use and even though we had a large data to process it was easily loaded and was available to use.