**Data Integration**

**Team Composition:**

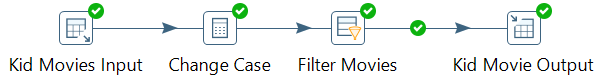
* Raghav Arora
* Kimberley Bhunu
* Greg Larmour
* Athena Kwok
* Max Sohl

**Transformation 1: Kid-Friendly Movies**

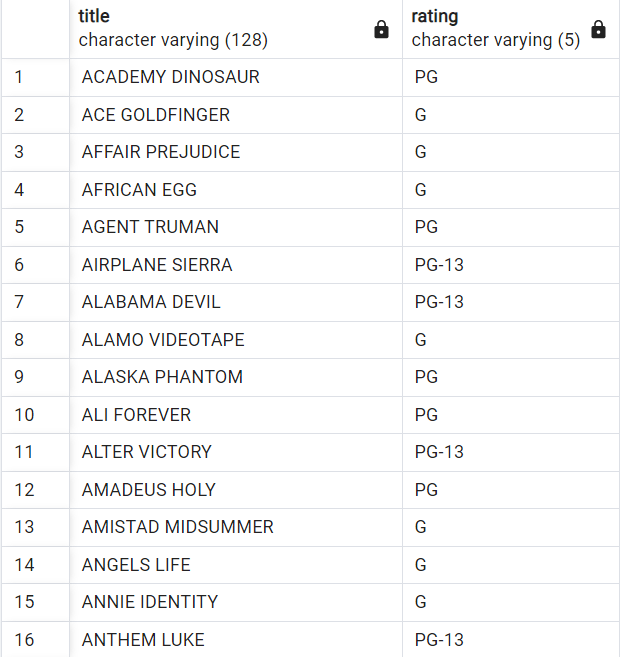
1. Describe:

The company is contracting with a local toy store to open a kiosk that targets children. The goal of this kiosk is to help increase DVD rental sales by providing the DVD rental store with a new location, thus a new stream of customers. In order to do so, a transformation is required that allows us to easily classify movies that are “G”, “PG”, and “PG13”. This is important as it will easily help segregate movies that are kid-friendly. On top of this, the managers decided they hated the “all caps lock” look of the movie titles within the database and would like the titles to look better when they copy and paste the values into important emails with the toy store. This will save valuable employee time in the fast-paced DVD rental game.

1. Assumptions and Rules:
   1. We assume the movies rated “G”, “PG” and “PG13” are all kid friendly.
   2. We assume opening a new kiosk will bring in a new stream of customers thereby increasing sales
   3. We assume that all the movies are already rated accurately.
   4. The transformation should include an input and output table which are considered to be steps
   5. The transformation will only show the movies that are kid-friendly
   6. The transformation process will use a filter function
   7. The transformation process will use a change case function to transform movie names to lowercase
2. Screenshot of Transformation Steps:



1. Show Results Table:

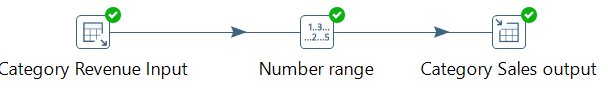


**Transformation 2: Categories With Highest Revenue**

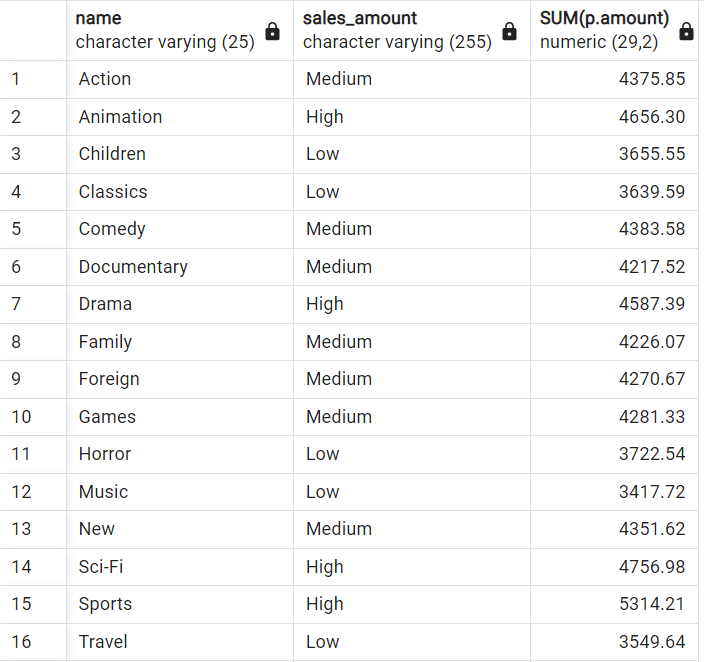
1. Describe:

The DVD rental store is trying to figure out which film categories bring in the highest revenue so they can increase inventory and advertising efforts/placement of those movies in order to increase rental sales. The transformation is required because it helps the DVD rental to classify film categories as low, medium, or high sales amounts. This is important as the output table will make it easier for the DVD store to distinguish sales for such categories.

1. Assumption and Rules:
   1. We assume that all the film categories have been categorized accurately beforehand
   2. We assume that the payment amount is directly related to revenue
   3. The transformation process will include an input and output table
   4. The transformation process will use a number range to categorize sales as low, medium, high
   5. The output table will show the aggregated payment amount (revenue) grouped by the film category
2. Screenshot of Transformation Steps:



1. Show Results Table:

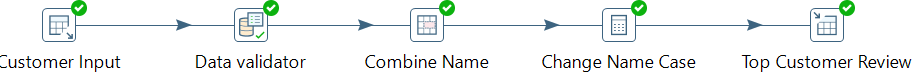


**Transformation 3: Top 100 Customer Reward Voucher**

1. Describe:

Our company heavily values our loyal repeat customers and has decided to provide our top 20 customers of the year with a voucher giving them 15% off future rentals for the next month. We will advertise this yearly reward system incentivizing other customers to rent more. The transformation is important for organizational reasons as it helps the DVD rental company to view the information more easily. The output table is important because it shows the store, which customers spent the most on DVDs and thus helps determine promotions.

1. Assumptions and Rules:
   1. We assume that customers that have the highest ‘amount’ in the payment table are the most frequent and most loyal users
   2. We assume that customers will strive to reach the top-20 rank in each store, thus increasing total sales
   3. We assume that the reward system wills strengthen customer loyalty for the highest-ranking customers
   4. The transformation process will include an input and output table
   5. The transformation process will include a data validator function to check if data types are accurate
   6. The transformation process will use the combine function to combine first names and last names
   7. The transformation process will use name cases to convert names to lowercase
2. Screenshot of Transformation Steps:



1. Show Results Table:

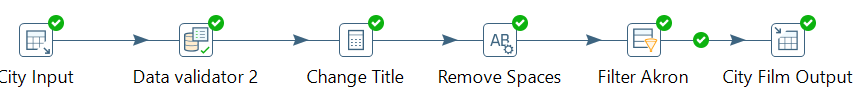


**Transformation 4: Movie Rentals in the Most Popular Rental City (Akron)**

1. Describe:

The company by running a quick SQL query has discovered that Akron is the home city of most of their customers. The company is trying to figure out which movie titles were rented in Akron. The company has decided to do this as they plan to hold a month-long festival in July where they stream a movie each night. Their logic is that doing so will further increase their brand recognition amongst customers in their most popular city. The transformation is required as it allows the company to filter by city and see which titles were rented in that city. The output table is important as it shows the DVD rental which movie titles should be played at the festival.

1. Assumptions and Rules:
   1. Assume the festival will bring in more future customers
   2. We can assume that the movie titles rented in Akron are present in our inventory.
   3. The transformation process will include an input and output table
   4. The transformation process will include a data validator function to check if data types are accurate
   5. The transformation process uses name case to change the case in the film title, remove spaces in the film title and city, and filters the data to only include the city of Akron, Ohio
2. Screenshot of Transformation Steps:



1. Show Results Table:



**Transformation 5: Staff Directory**

1. Describe:

Sakila has realized a massive problem in their entry of employee emails. InStaff email addresses should have been entered as all lowercase! What a blunder! They also realize the same person who made this mistake made an even worse error for the entire staff username list. After firing this terrible data entry employee, they now must fix these terrible errors. A daunting task indeed.

1. Assumptions and Rules:
   1. The emails were entered incorrectly.
   2. The updated emails match the employee’s actual email
   3. The new user name design will be more user-friendly for both staff members and management.
   4. The transformation process will include an input and output table
   5. The transformation process will include a data validator function to check if data types are accurate
   6. The transformation process uses a change case for username, password, last name, and email
   7. The transformation process uses removes spaces from the last name, and concatenates both password and last name
2. Screenshot of Transformation Steps:



1. Show Results Table:

