**Project 3: share the outputs from the operations you did using Knime and show the charts for both data sets. Show the visual of your Knime workflow. Explain the value of the particular operations you did. Summarize the key skills you developed.**

Data Story: The goal of the project was to analyze Yelp data to identify the most suitable city and cuisine type for setting up a new restaurant. Our workflow included multiple steps to process the data, derive meaningful insights, and visualize the results. We chose to focus on finding the top-rated restaurants, comparing cuisine types, and selecting a city for establishment and expansion.

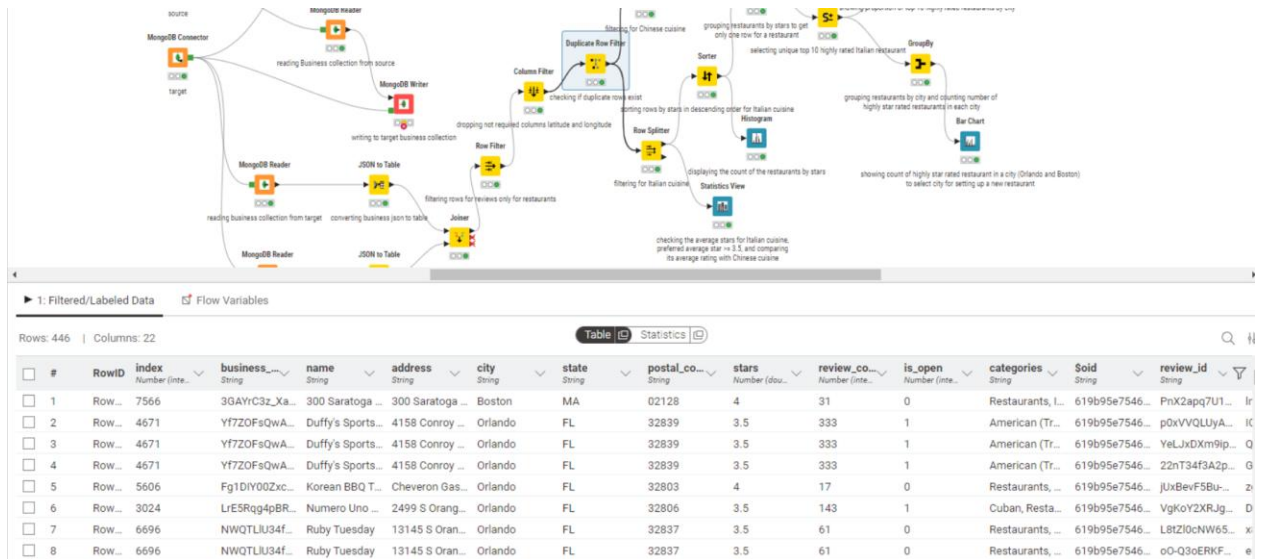1. **Outputs from the operations:**

a. Row Filter:
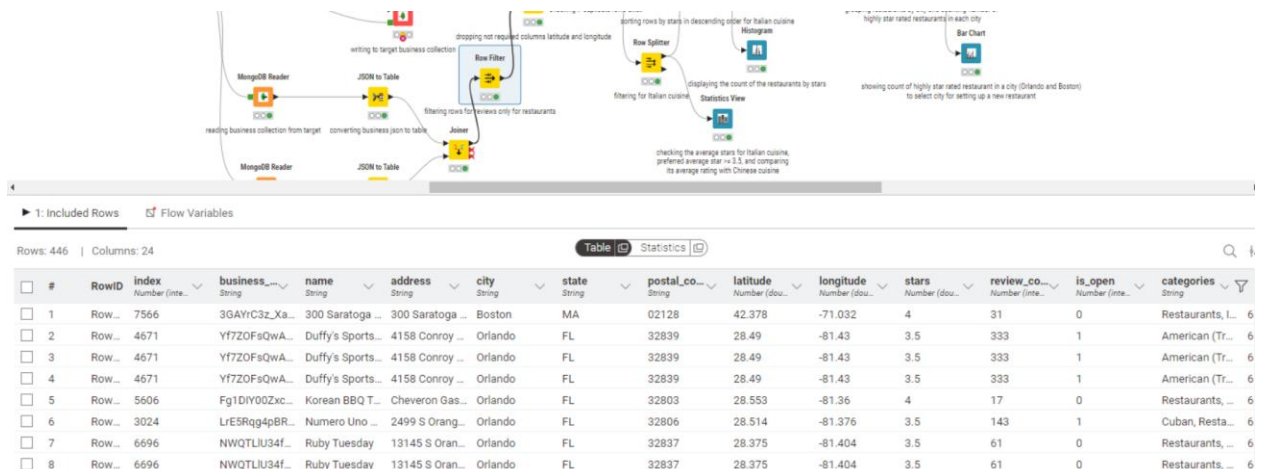
This operation filtered the joined tables of business and review of Yelp by rows to return observations for reviews only for "restaurants" category.
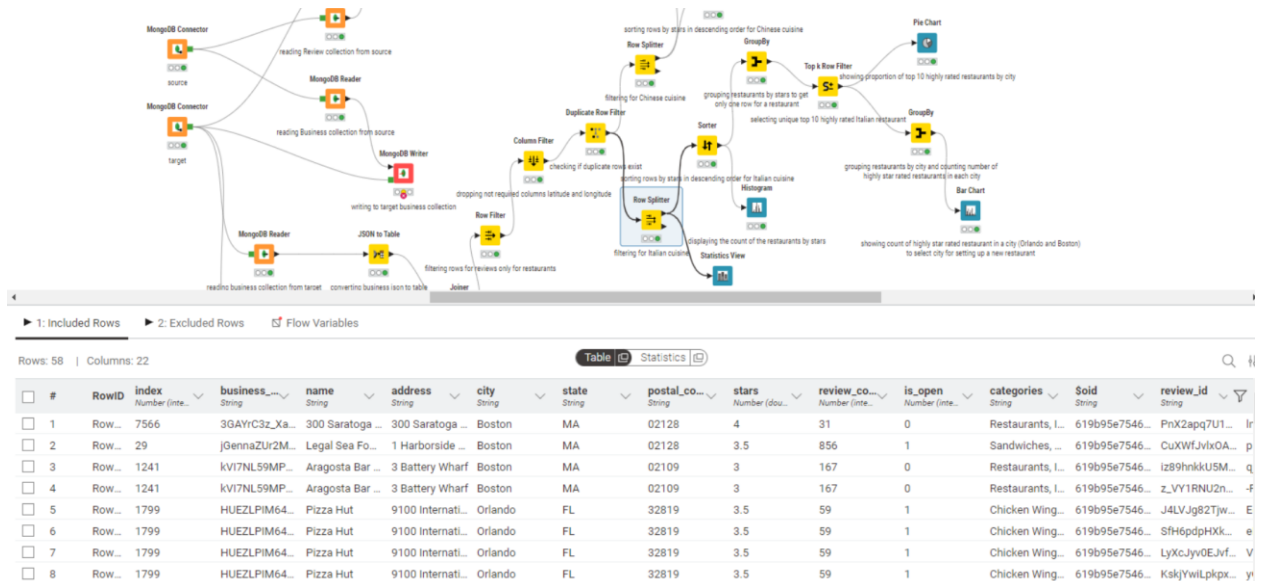
b. Column filter:

This operation aimed at dropping columns "latitude" and "longitude" that we felt were not required for our data story.

c. Duplicate Row Filter:

This operation aimed at ensuring there are no duplicate rows in the table.

d. Row Splitter:

This operation involved filtering rows only for Italian and Chinese cuisines separately. The two distinct operations were done to assess and compare Italian and Chinese restaurants and decide what type of restaurant to proceed with based on ratings.

e. Sorter:
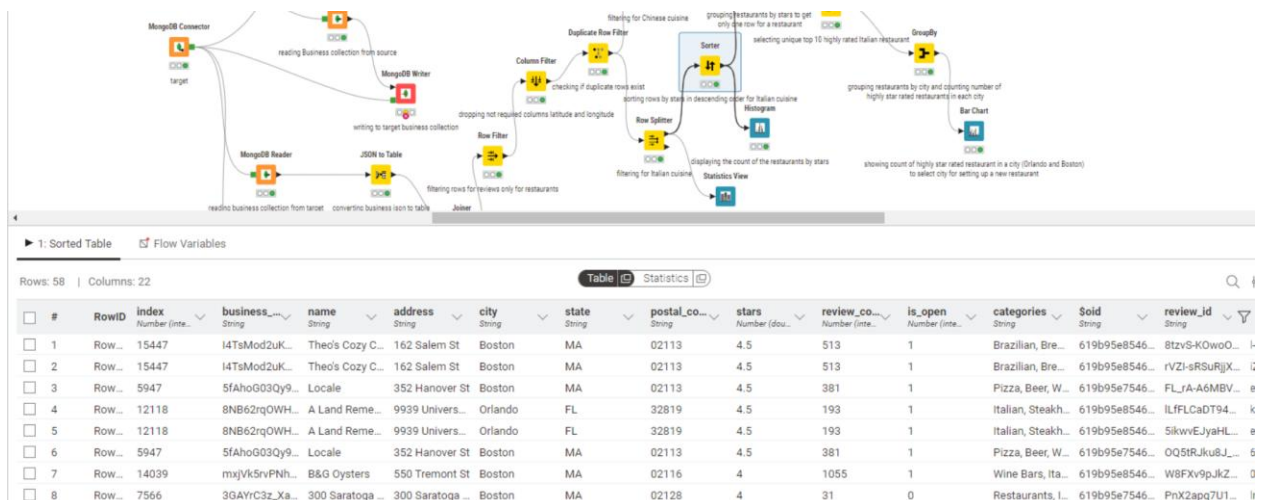
This operation aimed at sorting rows for both Italian and Chinese restaurants tables by "stars" in a descending order, in order to get the highly rated restaurants at the top in our tables.

f. Group By:

| #	RowID	business_id String	city String	state String	stars Set
1	Row9	I4TsMod2uKwI1ftm0fd2hQ	Boston	MA	[4.5]
2	Row3	5fAhoG03Qy99Ii0v7jGFYg	Boston	MA	[4.5]
3	Row5	8NB62rqOWHy3zDck4cCBhQ	Orlando	FL	[4.5]
4	Row24	mxjVk5rvPNhzYe_vt3OSQA	Boston	MA	[4.0]
5	Row1	3GAYrC3z_XaNP2k8qJP-lg	Boston	MA	[4.0]
6	Row0	1VqftFBUMdktbt2MX5JUNQ	Boston	MA	[4.0]
7	Row11	J3Xb6uihPkZOKe4wNmYGFQ	Boston	MA	[4.0]
8	Row15	QFEEm3joTWU5ZF4suCuRPQ	Boston	MA	[4.0]
9	Row13	MjpH-uP90jTUp_KqAOiBJg	Boston	MA	[4.0]

We performed a group by operation at this stage to group the rows by stars and get only one row for a particular business/restaurant. This will ensure that for the next steps, we get to the top 10 unique restaurants in our result. Without this operation, if we select the top 10 rows, there is possibility that we only get 7 unique restaurants because of multiple rows for a restaurant. As a result of this, instead of the top 10, we will only get top 7 restaurants.

g. Top k Row Filter:
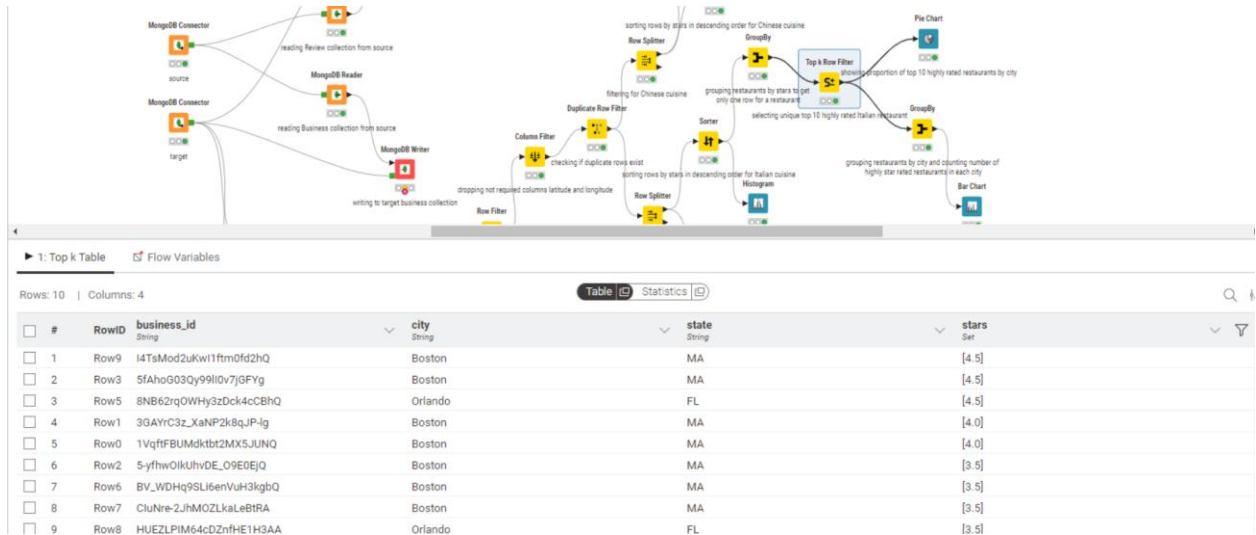
This operation selects the top 10 unique restaurants from our sorted rows (by stars) in the table.
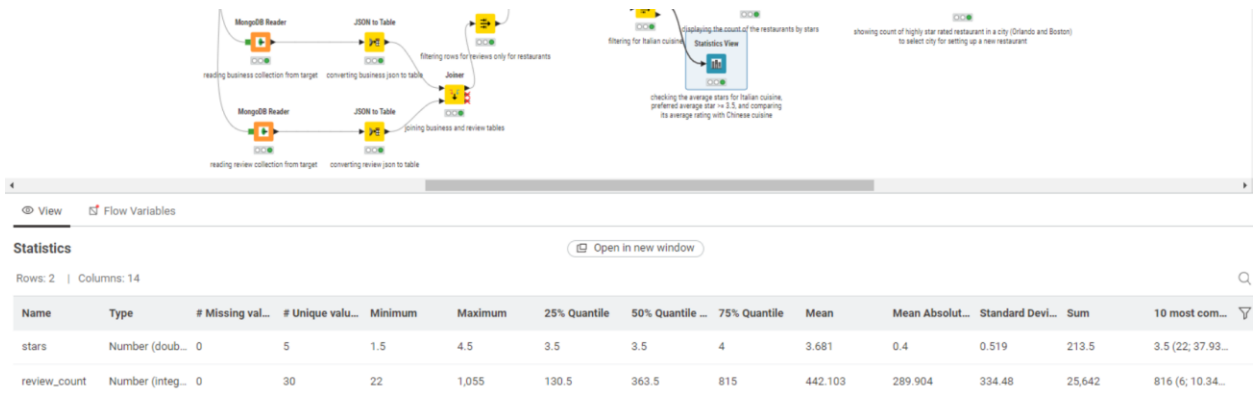
h. Group By:
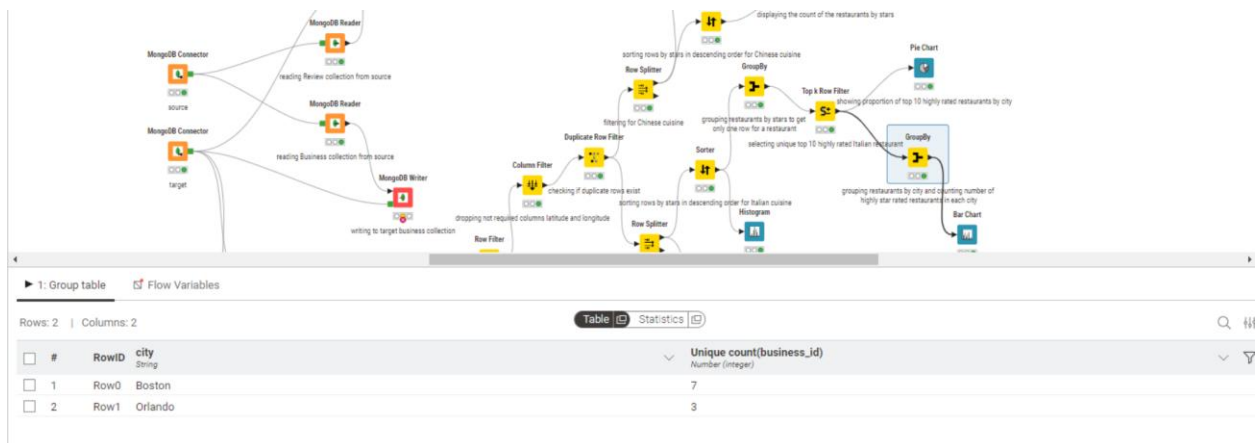
We again performed group by operation here to group the restaurants by city and count the total number of highly rated Italian restaurants in each city. This was done to analyze which city has the highest number of highly rated Italian restaurants so that we can assess which city would be preferable to set up our new Italian restaurant. Based on the counts and the visualizations, we decided to select Orlando housing 3 out of 7 top 10 highly rated

restaurants, as we feel that our business will have fewer competitors in Orlando compared to Boston and hence, better opportunities for our new business to profit and grow.

2. Charts:

a. Statistics View:

Statistics
Rows: 2 | Columns: 14

| Name	Type	# Missing val...	# Unique valu...	Minimum	Maximum	25% Quantile	50% Quantile ...	75% Quantile	Mean	Mean Absolut...	Standard Devi...	Sum	10 most com...
stars	Number (doub...	0	5	1.5	4.5	3.5	3.5	4	3.681	0.4	0.519	213.5	3.5 (22; 37.93...
review_count	Number (integ...	0	30	22	1,055	130.5	363.5	815	442.103	289.904	334.48	25,642	816 (6; 10.34...

This summary statistics for both Italian and Chinese provides us with insights on the average rating for both types of cuisines so we can decide which of the two types has decent average rating and which one can be used while setting up a new restaurant. Even with higher number of columns, we get better average rating for Italian type compared to Chinese type which indicates the people prefer Italian cuisines more than Chinese ones. Hence, we decided to open an Italian restaurant in a particular city.

b. Histogram:

Count of Restaurants by Stars

We counted and displayed Italian and Chinese restaurants by star ratings. We discovered that Italian cuisine has restaurants lying mostly in the range of 3-4.5 stars ratings, while Chinese cuisines have restaurants mostly lying in the 3.3.2 and 3.8-4 (less compared to Italian 13 vs 27). This analysis further solidified our decision to choose an Italian restaurant to set up over a Chinese restaurant.

c. Pie Chart:


Proportion of Top 10 Restaurants by City
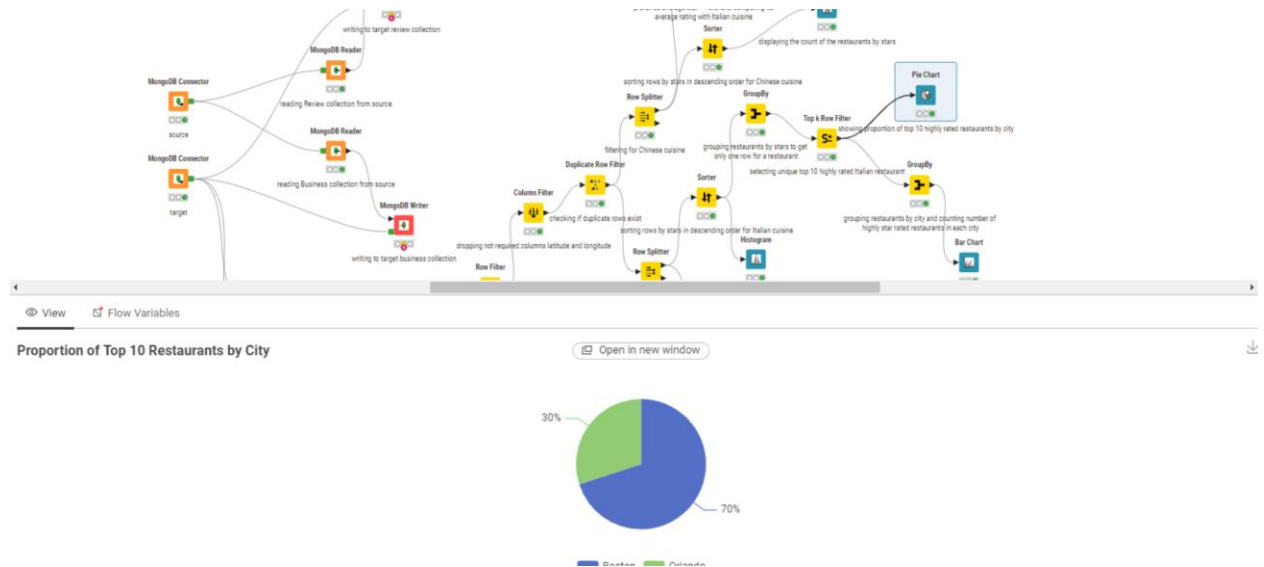
This pie chart shows the proportion of top 10 highly rated restaurants by city helping us analyze which city houses top 10 restaurants and if that city can be selected or not to set up our new Italian restaurant business.

d. Bar Chart:

Count of Highly Rated Restaurants in a City

This graph gave us a clear understanding on how the division of top 10 Italian restaurants looks like across Orlando and Boston. Based on the highest count of 7 out of 10 restaurants in Boston, we decided to select Orlando as our location to set up our new Italian restaurants because of fewer competition and better opportunities.

3. **Visual**:
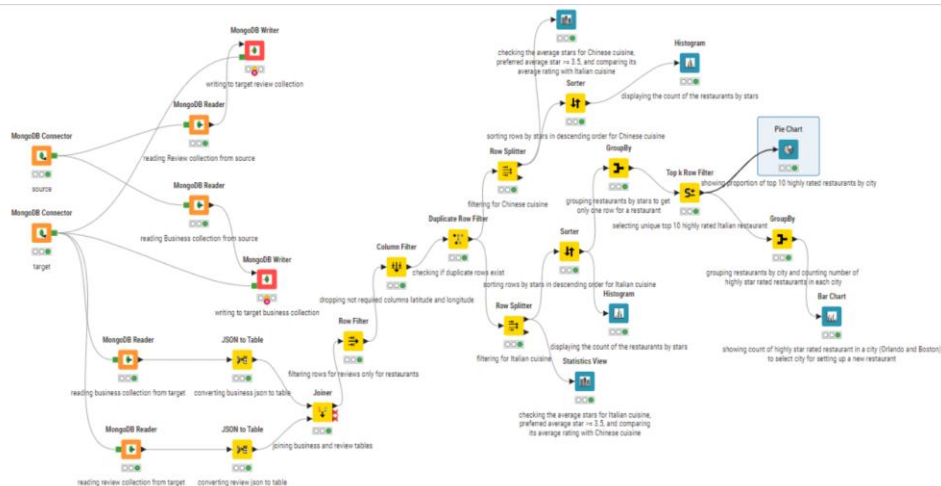
4. **Skills Developed:**

Through this project, we honed advanced data analysis and transformation skills, gaining proficiency in preparing, cleaning, and structuring datasets for meaningful insights. Leveraging tools like KNIME, we mastered the use of nodes such as Row Filter, Column Filter, GroupBy, and Top K Selector to handle complex workflows efficiently. Our ability to create impactful visualizations—bar charts, pie charts, and more—demonstrated a strong grasp of storytelling through data, ensuring clarity and relevance in decision-making. Additionally, we developed a strategic mindset by aligning technical operations with business objectives, showcasing our ability to turn raw data into actionable insights. These skills collectively enhanced our technical expertise and ability to present data-driven recommendations effectively.

5. **Finally, share one thing that you found most challenging and one thing you found most rewarding across all the projects:**

Most Challenging:
Challenge: Ensuring a cohesive narrative while performing complex operations.
Solution: Restructured the workflow two times to focus on answering one overarching question clearly and cohesively, which further clarified the steps and outputs.

Most Rewarding: The ability to make a confident, data-backed decision about the restaurant's location and cuisine type, which felt impactful and actionable.