Enterprise Database Management - Group 2

What we have learned and gained from three projects?

Group member:
Lechen Gong
Tzu-Ching Huang
Kumari Neha Priya
Jiahe Li

Project 1

```
query = """
SELECT TOP 5
    categories,
    COUNT(business_id) AS num_restaurants
FROM BUSINESS
WHERE city = 'Philadelphia' AND categories LIKE '%Restaurants%'
GROUP BY categories
ORDER BY num_restaurants DESC;
"""

# Execute the query and store the results in a DataFrame
df_popular_categories = pd.read_sql(query, engine)
print(df_popular_categories)
```

```
              categories  num_restaurants
0    Restaurants, Pizza               10
1  Restaurants, Chinese                5
2    Pizza, Restaurants                4
3   Korean, Restaurants                2
4  Mexican, Restaurants                2
```

Insights from the query:

- understand the competitive landscape and cuisine diversity in Philadelphia
- balance popularity and competition level
- the pizza and Chinese restaurants dominate the market
- may choose to set up Korean or Mexican restaurants

Project 1

```
query = """
SELECT categories, AVG(stars) as avg_rating
FROM BUSINESS
WHERE city = 'Philadelphia' AND categories LIKE '%Restaurants%'
GROUP BY categories
ORDER BY avg_rating DESC;
"""

# Execute the query and store the results in a DataFrame
df_rate = pd.read_sql(query, engine)
print(df_rate)
```

```
                                            categories  avg_rating
0    Food, Food Trucks, Mediterranean, Sandwiches, ...         5.0
1                                French, Restaurants           5.0
2    Food, Seafood, Sardinian, Gelato, Restaurants,...         4.5
3    Food, Juice Bars & Smoothies, Coffee & Tea, Re...         4.5
4    Grocery, Vegan, Food, Vegetarian, Convenience ...         4.5
..                                                 ...         ...
175  Restaurants, Fast Food, Food, Coffee & Tea, Bu...         1.5
176         Italian, Chicken Wings, Restaurants, Pizza         1.5
177                    Burgers, Restaurants, Fast Food         1.5
178  Fast Food, Restaurants, Food, Coffee & Tea, Bu...         1.0
179  Fast Food, Food, Burgers, Coffee & Tea, Restau...         1.0

[180 rows x 2 columns]
```

Insights from the query:

- understand which cuisines have strong customer appreciation
- higher rating for Mediterranean and French cuisines
- may choose to specialize in Mediterranean or French cuisine, and sell a variety of drinks

Project 1

```
query = """
SELECT TOP 5 name, review_count
FROM BUSINESS
WHERE city = 'Philadelphia' AND categories LIKE '%Restaurants%'
ORDER BY review_count DESC;
"""

# Execute the query and store the results in a DataFrame
df_mostreview = pd.read_sql(query, engine)
print(df_mostreview)
```

```
                        name  review_count
0                 Han Dynasty           783
1                     In Riva           729
2  Bareburger - Midtown Village           622
3                    The Love           618
4                Pub & Kitchen           615
```

Insights from the query:

- top 5 restaurants in Philadelphia with the most customer reviews
- learn their strengths and improve our own restaurants
- help us avoid these problems in future restaurant operations

Project 1

```
query = """
SELECT review_id, text
FROM REVIEW
WHERE business_id IN (SELECT business_id FROM BUSINESS WHERE city = 'Philadelphia')
AND text LIKE '%service%';
"""

# Execute the query and store the results in a DataFrame
df_feedback = pd.read_sql(query, engine)
print(df_feedback)
```

```
                review_id                                               text
0   NABhrXkQZp4hhCg_47odVA  Mind blown. This is the best shop I have ever ...
1   VuNZ2IbgXge-X-HsTuzCbA  Great cheesesteak! Friendly and fast service. ...
2   oVSmMGam9q_oh2XMNjDCcw  This is our go-to takeout place. It has great ...
3   JR6yMkNFaYz8ZwxnS2IJNg  I want to thank Eric @paintlessautodents.com \...
4   beE-8wCIxCzVIrsgQgahwQ  Terrible service, food came out cold even afte...
..                     ...                                                ...
65  EeIebnG4LxdcAyWIkJNVJg  Not a vegan or a vegetarian but I love eating ...
66  mfw86JhavKYEDLAcAbw6EA  Delicious and very nicely presented. We had J ...
67  cRHR6glpwKI93bnSlszrgw  Horrible customer service for overpriced drink...
68  ub2BMOwKSTbhbuovib8Gtw  I feel so comfortable with the Mazzoni Center....
69  pbXbIoZwe6cEihaYqbkSHw  It was a pretty cute little restaurant; the at...

[70 rows x 2 columns]
```

Insights from the query:

- list all customer reviews for restaurants services in Philadelphia
- help us to identify what customers value most in restaurant service
- use these review feedback to shape our hiring, training, and customer service policies in our restaurant

Project 1

```
query = """
SELECT attributes, COUNT(*) as frequency
FROM BUSINESS
WHERE city = 'Philadelphia' AND stars >= 4
GROUP BY attributes
ORDER BY frequency DESC;
"""

# Execute the query and store the results in a DataFrame
df_attributes = pd.read_sql(query, engine)
print(df_attributes)
```

```
                                            attributes  frequency
0                                                 None         22
1                {'BusinessAcceptsCreditCards': 'False'}          4
2    {'BusinessAcceptsCreditCards': 'True', 'ByAppo...          4
3                 {'BusinessAcceptsCreditCards': 'True'}          3
4                               {'GoodForKids': 'True'}          2
..                                                  ...        ...
197  {'BusinessAcceptsCreditCards': 'False', 'Busin...          1
198  {'BusinessAcceptsCreditCards': 'False', 'Busin...          1
199  {'BusinessAcceptsCreditCards': 'False', 'ByApp...          1
200  {'BusinessAcceptsCreditCards': 'False', 'ByApp...          1
201  {'BusinessAcceptsCreditCards': 'False', 'ByApp...          1

[202 rows x 2 columns]
```

Insights from the query:

- understand which elements resonate most with customers
- incorporate the top-rated attributes into our restaurant concept

Key Skills For Project 1

- better understand of advanced SQL techniques
- cultivate our business insights by analyzing Yelp data to build our own restaurant
- analyze the saturation of the restaurant market, the advantages and characteristics of high-scoring restaurants, and the reasons for high customer loyalty
- better prepare our future projects and careers

Project 2

Which token we picked and why?

- an experimental meme token created to build a unique culture supported by cryptocurrency members
- explore the performance outcome of this experimental meme coin
- understand the transaction pattern

Persona: Venture Capitalists

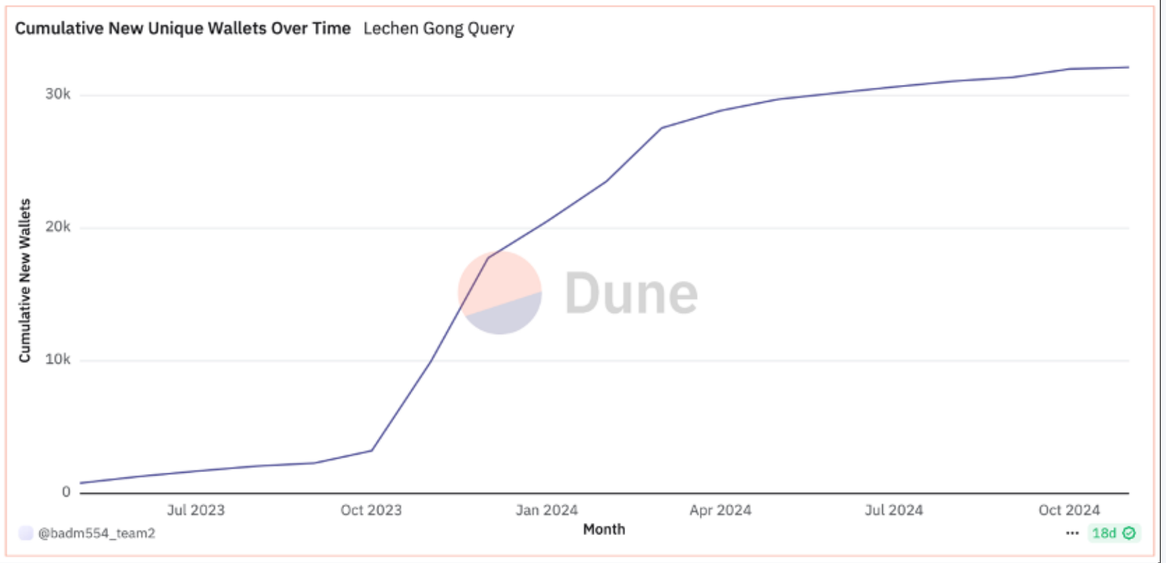
Intent:

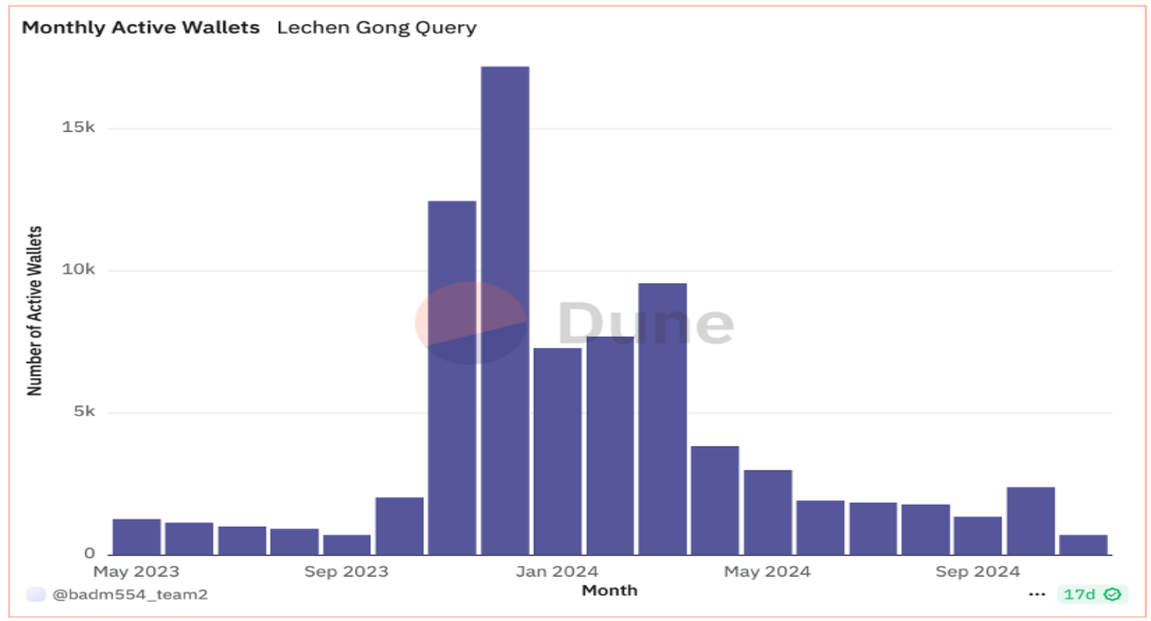
- track WOOF’s community growth
- understand transaction trends and assess whale impacts (centralization risk)
- help the decision-making process on WOOF’s investment potential

Two key concepts about blockchains

- **Blockchain structure and conten**t
  - fundamental units of a blockchain, containing transaction data and essential metadata
  - each block typically includes block header, transaction data and block size
  - ensure data integrity and immutability
- **Key values to analysis**
  - utilize data processing techniques to examine transaction histories, wallet addresses, and smart contract interactions
  - detect suspicious activities, ensure compliance, and maintain blockchain network integrity
  - ensure transparency and trust

Project 2



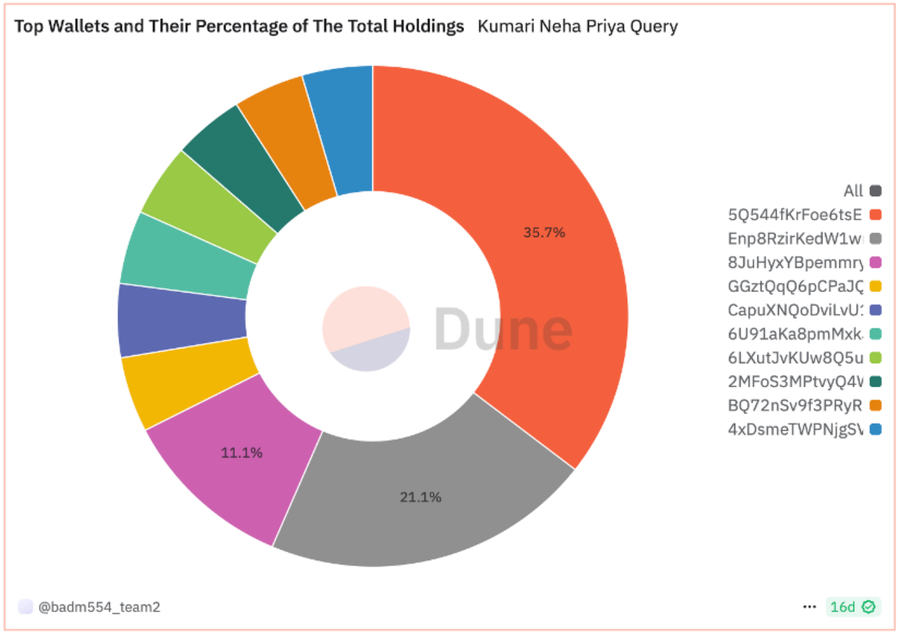


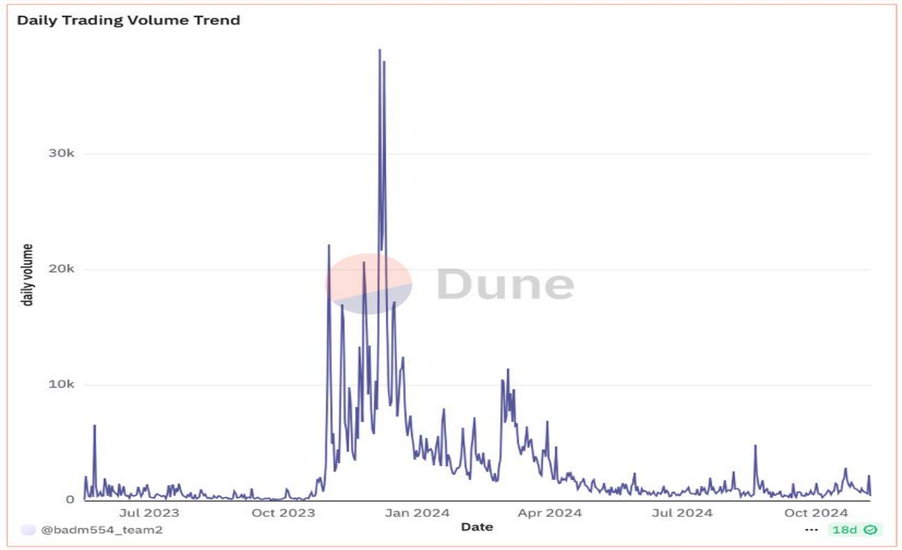
Insights from the query:

- assess WOOF’s sustainability by examining community growth
- WOOF has expanding user base over time and the gradual flattening trend suggests sustainable growth

- track monthly active users trading the WOOF
- suggest less and less users interested in trading WOOF

Project 2

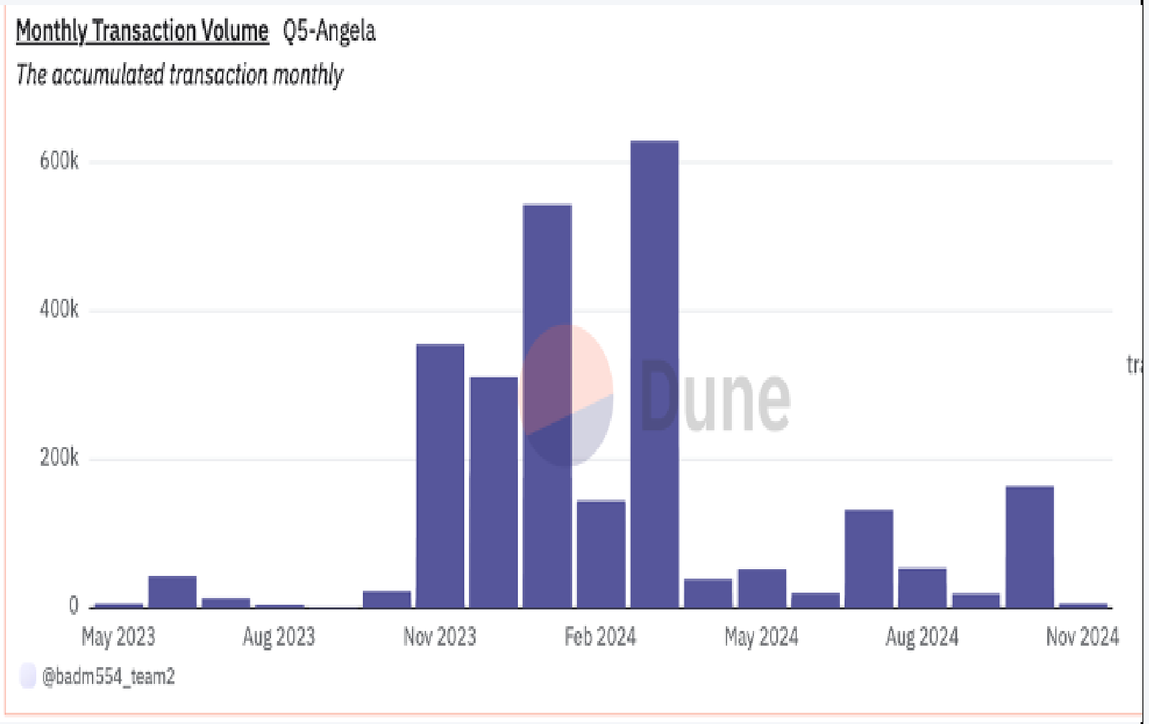




Insights from the query:

- assess whale concentration and centralization risk
- indicate potential price manipulation risks if these whales decide to sell large amounts
- identify top holders, assessing centralization risks and potential market manipulation by large token holders
- indicate whale activity or promotional events
- consistent low trading volume in mid to late 2024 suggests a smaller but steady user base

Project 2



Insights from the query:

- assess whale concentration and centralization risk
- understand the distribution of activity among whales
- provide insights into the concentration of trading activity and its dynamics over time
- indicate changes in trading behavior
- suggest vulnerability to market volatility
- inform strategies to attract new participants or retain high-volume traders

Key Skills For Project 2

- **Blockchain Analysis:**
  - understand blockchain structure and analyze transaction histories, wallet addresses, and smart contract interactions to detect suspicious activities and ensure compliance
- **Data Interpretation:**
  - use data queries and visualizations to identify trends in user growth, activity, and whale concentration, and interprete their implications for token sustainability and risk
- **Strategic Assess:**
  - evaluate investment potential by assessing community growth, centralization risks, and trading activity patterns to inform decision-making processes.

Project 3

▶ 1: Included Rows Flow Variables

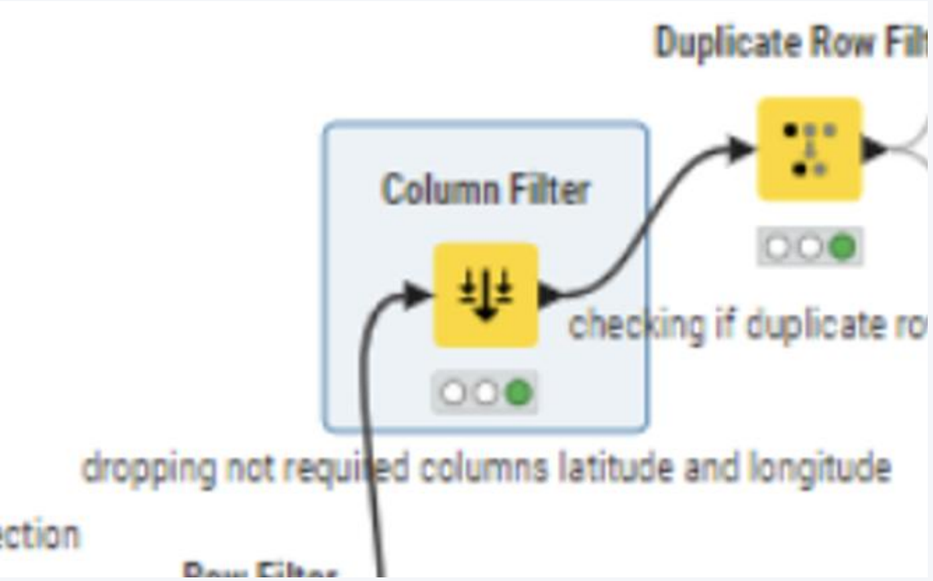
Rows: 446 | Columns: 24

#	RowID	index *Number (inte...*	business_... *String*	name *String*	address *String*	city *String*
1	Row...	7566	3GAYrC3z_Xa...	300 Saratoga ...	300 Saratoga ...	Boston
2	Row...	4671	Yf7ZOFsQwA...	Duffy's Sports...	4158 Conroy ...	Orlando
3	Row...	4671	Yf7ZOFsQwA...	Duffy's Sports...	4158 Conroy ...	Orlando
4	Row...	4671	Yf7ZOFsQwA...	Duffy's Sports...	4158 Conroy ...	Orlando
5	Row...	5606	Fg1DIY00Zxc...	Korean BBQ T...	Cheveron Gas...	Orlando
6	Row...	3024	LrE5Rqg4pBR...	Numero Uno ...	2499 S Orang...	Orlando
7	Row...	6696	NWQTLlU34f...	Ruby Tuesday	13145 S Oran...	Orlando
8	Row...	6696	NWQTLlU34f...	Ruby Tuesday	13145 S Oran...	Orlando

Insights:

- filter the joined tables of business and review of Yelp by rows to return observations for reviews only for “restaurants” category

Project 3



▶ 1: Filtered table | Flow Variables

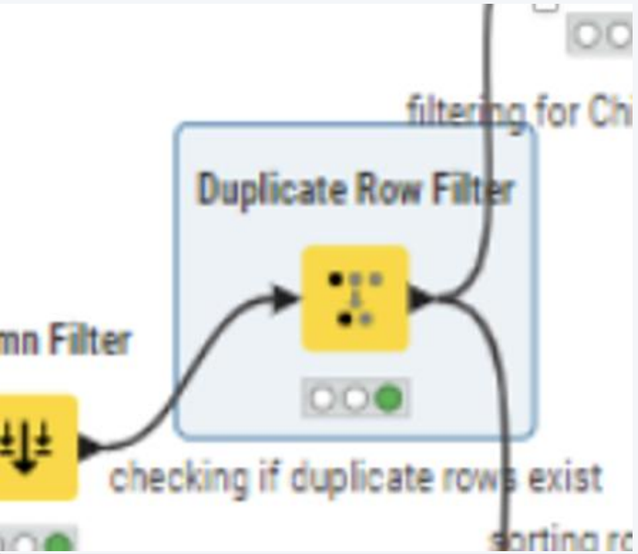
Rows: 446 | Columns: 22

#	RowID	index *Number (inte...*	business_... *String*	name *String*	address *String*	city *String*
1	Row...	7566	3GAYrC3z_Xa...	300 Saratoga ...	300 Saratoga ...	Boston
2	Row...	4671	Yf7ZOFsQwA...	Duffy's Sports...	4158 Conroy ...	Orlando
3	Row...	4671	Yf7ZOFsQwA...	Duffy's Sports...	4158 Conroy ...	Orlando
4	Row...	4671	Yf7ZOFsQwA...	Duffy's Sports...	4158 Conroy ...	Orlando
5	Row...	5606	Fg1DIY00Zxc...	Korean BBQ T...	Cheveron Gas...	Orlando
6	Row...	3024	LrE5Rqg4pBR...	Numero Uno ...	2499 S Orang...	Orlando
7	Row...	6696	NWQTLlU34f...	Ruby Tuesday	13145 S Oran...	Orlando
8	Row...	6696	NWQTLlU34f...	Ruby Tuesday	13145 S Oran...	Orlando

Insights:

- aim at dropping columns “latitude” and “longitude” that we felt were not required for our data story

Project 3



▶ 1: Filtered/Labeled Data Flow Variables

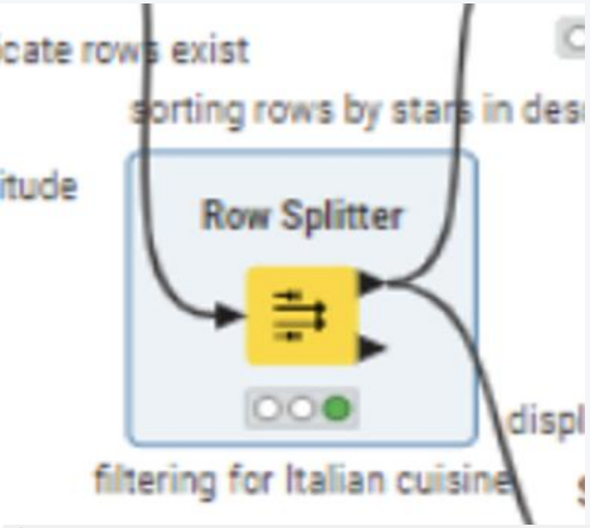
Rows: 446 | Columns: 22

#	RowID	index *Number (inte...*	business_... *String*	name *String*	address *String*	city *String*
1	Row...	7566	3GAYrC3z_Xa...	300 Saratoga ...	300 Saratoga ...	Boston
2	Row...	4671	Yf7ZOFsQwA...	Duffy's Sports...	4158 Conroy ...	Orlando
3	Row...	4671	Yf7ZOFsQwA...	Duffy's Sports...	4158 Conroy ...	Orlando
4	Row...	4671	Yf7ZOFsQwA...	Duffy's Sports...	4158 Conroy ...	Orlando
5	Row...	5606	Fg1DIY00Zxc...	Korean BBQ T...	Cheveron Gas...	Orlando
6	Row...	3024	LrE5Rqg4pBR...	Numero Uno ...	2499 S Orang...	Orlando
7	Row...	6696	NWQTLlU34f...	Ruby Tuesday	13145 S Oran...	Orlando
8	Row...	6696	NWQTLlU34f...	Ruby Tuesday	13145 S Oran...	Orlando

Insights:

- aim at ensuring there are no duplicate rows in the table

Project 3



▶ 1: Included Rows ▶ 2: Excluded Rows Flow Variables

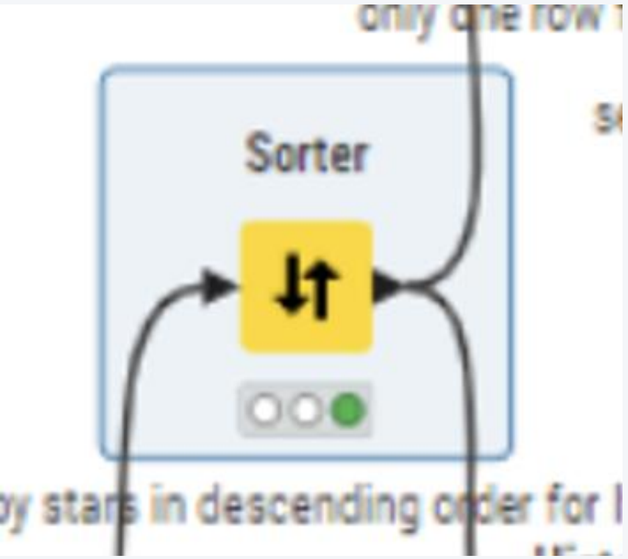
Rows: 58 | Columns: 22

#	RowID	index *Number (inte...*	business_... *String*	name *String*	address *String*	city *String*
1	Row...	7566	3GAYrC3z_Xa...	300 Saratoga ...	300 Saratoga ...	Boston
2	Row...	29	jGennaZUr2M...	Legal Sea Fo...	1 Harborside ...	Boston
3	Row...	1241	kVI7NL59MP...	Aragosta Bar ...	3 Battery Wharf	Boston
4	Row...	1241	kVI7NL59MP...	Aragosta Bar ...	3 Battery Wharf	Boston
5	Row...	1799	HUEZLPIM64...	Pizza Hut	9100 Internati...	Orlando
6	Row...	1799	HUEZLPIM64...	Pizza Hut	9100 Internati...	Orlando
7	Row...	1799	HUEZLPIM64...	Pizza Hut	9100 Internati...	Orlando
8	Row...	1799	HUEZLPIM64...	Pizza Hut	9100 Internati...	Orlando

Insights:

- involve filtering rows only for Italian and Chinese cuisines separately
- 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

Project 3



Rows: 58 | Columns: 22

#	RowID	index *Number (inte...*	business_... *String*	name *String*	address *String*	city *String*
1	Row...	15447	I4TsMod2uK...	Theo's Cozy C...	162 Salem St	Boston
2	Row...	15447	I4TsMod2uK...	Theo's Cozy C...	162 Salem St	Boston
3	Row...	5947	5fAhoG03Qy9...	Locale	352 Hanover St	Boston
4	Row...	12118	8NB62rqOWH...	A Land Reme...	9939 Univers...	Orlando
5	Row...	12118	8NB62rqOWH...	A Land Reme...	9939 Univers...	Orlando
6	Row...	5947	5fAhoG03Qy9...	Locale	352 Hanover St	Boston
7	Row...	14039	mxjVk5rvPNh...	B&G Oysters	550 Tremont St	Boston
8	Row...	7566	3GAYrC3z_Xa...	300 Saratoga ...	300 Saratoga ...	Boston

Insights:

- aim at sorting rows for both Italian and Chinese restaurants tables by “stars” in a descending order
- get the highly rated restaurants at the top in our tables

Project 3

Rows: 30 | Columns: 4

#	RowID	business_id *String*	city *String*
1	Row9	I4TsMod2uKwI1ftm0fd2hQ	Boston
2	Row3	5fAhoG03Qy99lI0v7jGFYg	Boston
3	Row5	8NB62rqOWHy3zDck4cCBhQ	Orlando
4	Row24	mxjVk5rvPNhzYe_vt3OSQA	Boston
5	Row1	3GAYrC3z_XaNP2k8qJP-lg	Boston
6	Row0	1VqftFBUMdktbt2MX5JUNQ	Boston
7	Row11	J3Xb6uihPkZOKe4wNmYGFQ	Boston
8	Row15	QFEEm3joTWU5ZF4suCuRPQ	Boston
9	Row13	MjpH-uP90jTUp_KqAOiBJg	Boston

Insights:

- perform a group by operation at this stage to group the rows by stars and get only one row for a particular business/restaurant
- 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

Project 3

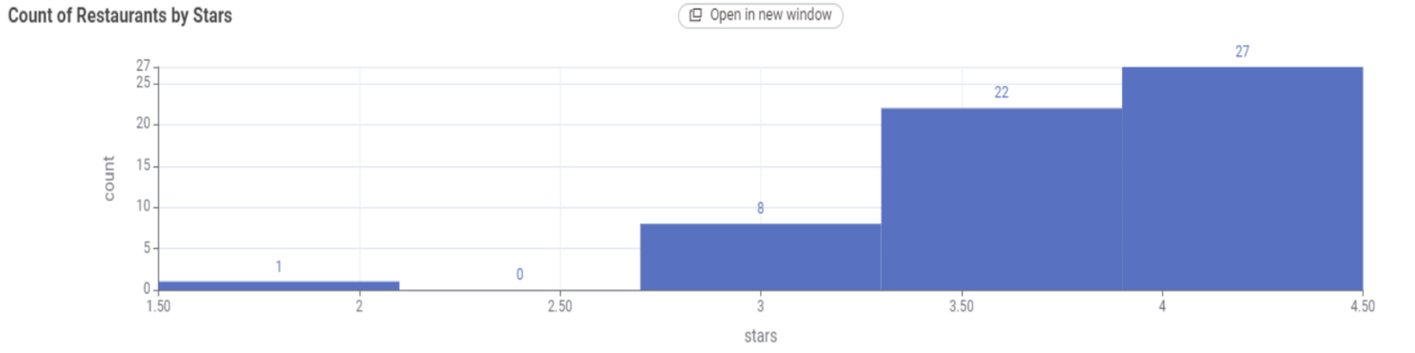
Rows: 10 | Columns: 4

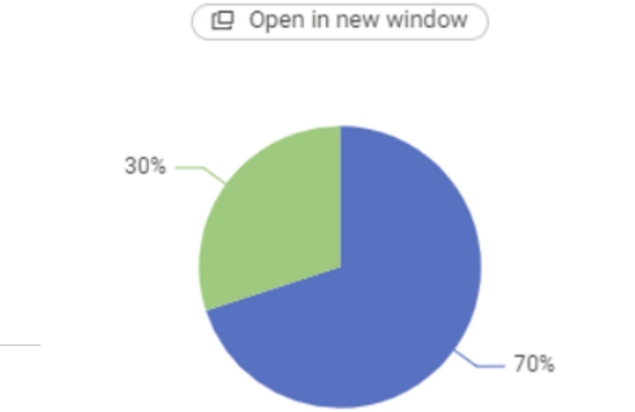
#	RowID	business_id *String*	city *String*
1	Row9	I4TsMod2uKwI1ftm0fd2hQ	Boston
2	Row3	5fAhoG03Qy99lI0v7jGFYg	Boston
3	Row5	8NB62rqOWHy3zDck4cCBhQ	Orlando
4	Row1	3GAYrC3z_XaNP2k8qJP-lg	Boston
5	Row0	1VqftFBUMdktbt2MX5JUNQ	Boston
6	Row2	5-yfhwOIkUhvDE_O9E0EjQ	Boston
7	Row6	BV_WDHq9SLi6enVuH3kgbQ	Boston
8	Row7	CIuNre-2JhMOZLkaLeBtRA	Boston
9	Row8	HUEZLPIM64cDZnfHE1H3AA	Orlando

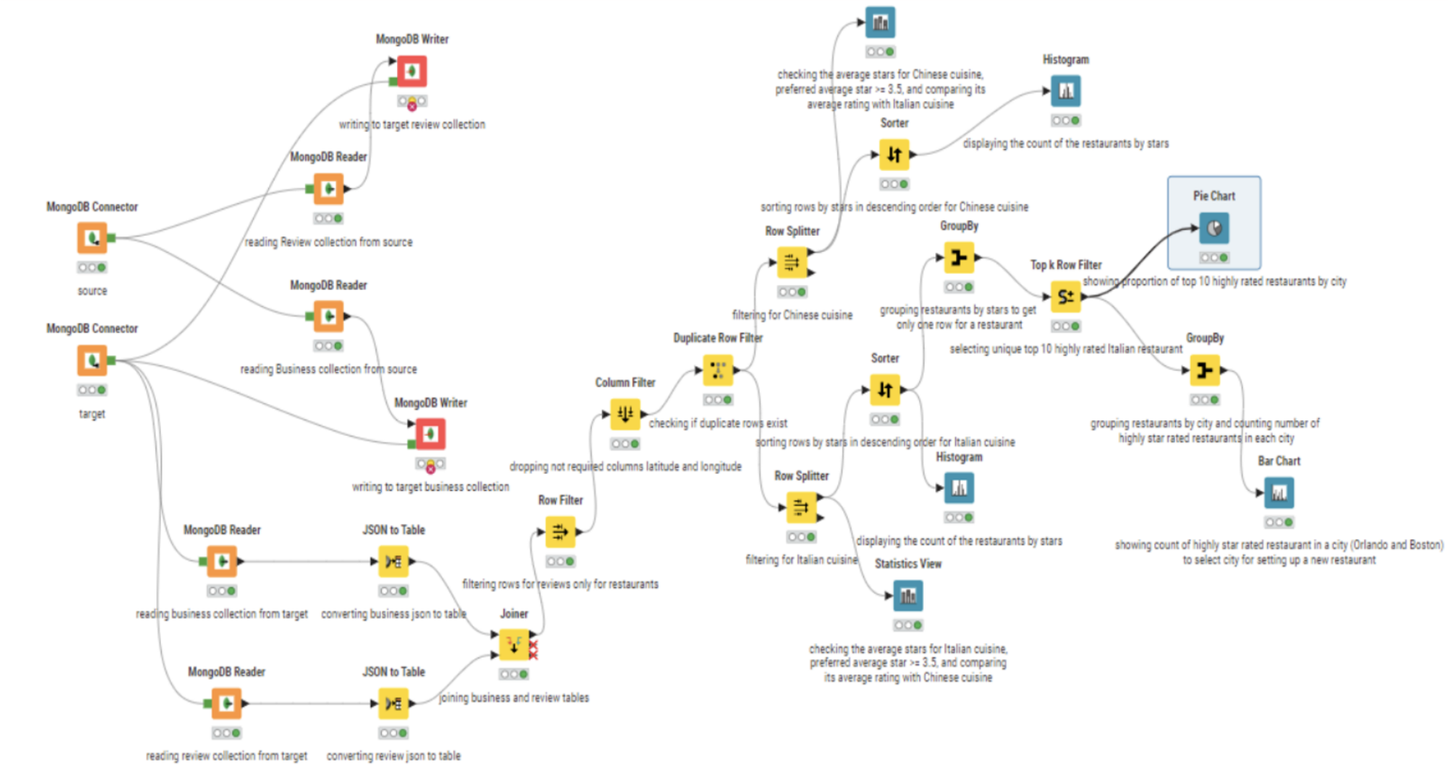
Insights:

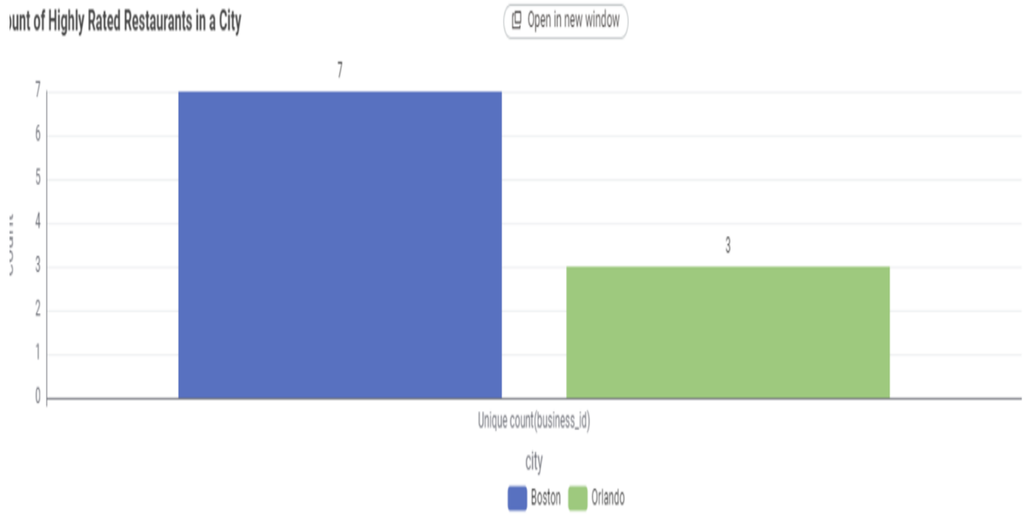
- select the top 10 unique restaurants from our sorted rows (by stars) in the table
- perform group by operation here to group the restaurants by city and count the total number of highly rated Italian restaurants in each city
- analyze which city has the highest number of highly rated Italian restaurants
- assess which city would be preferable to set up our new Italian restaurant.
- select Orlando housing 3 out of 7 top 10 highly rated restaurants

Project 3 Visual









Key Skills For Project 3

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

Challenge: Ensuring a cohesive narrative while performing complex operations

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

Most Challenging Part

When an error occurs but can’t be solved:

- same Dune query for selecting, same Python code for connecting, same nodes in Knime, missing values still exist

View | Flow Variables

Statistics | Open in new window

Rows: 27 | Columns: 14

Name	Type	# Missing val...	# Unique valu...	Minimum	Maximum	25% Quantile	50% Quantile ...	75% Quantile	Mean
blockchain	String	0	1						
fee_tier	Number (dou...	15	1	0.01	0.01	0.01	0.01	0.01	0.01
fee_usd	Number (dou...	15	2418	0	48.139	0.027	0.029	0.031	0.233
outer_instruct...	Number (inte...	0	9	1	11	3	4	4	3.554
project	String	0	2						
project_main_...	String	0	2						
project_progr...	String	3	13						

Most Rewarding Part

- **Practical Skills:**
  - put theoretical knowledge into practice(sql, python, cloud database, knime)
  - increase the possibility of finding a job
- **Patience:**
  - Patience is key in life

Thank You