

SNOWFLAKE-ASSIGNMENT

Dataset Selection

- The data is about the applications available on the App Store, with their current version, the rating they received, price and genre of application and number of supported devices of each application.
- Dataset Link: [App-Store-Dataset](#)

Data Loading

- Create a DataWarehouse named Rithvik_Warehouse.

```
1 CREATE WAREHOUSE IF NOT EXISTS Rithvik_Warehouse
2   WITH WAREHOUSE_SIZE = 'XSMALL'
3   AUTO_SUSPEND = 300
4   AUTO_RESUME = TRUE
5   INITIALLY_SUSPENDED = TRUE;
6
```

- Create a database named rithvik18_database

```
8
9 CREATE DATABASE IF NOT EXISTS rithvik18_database;
10
11
12
```

↶ Results ~ Chart

	status
1	Database RITHVIK18_DATABASE successfully created.

- Create a table named app_store_data

```
15 CREATE OR REPLACE TABLE app_store_data (  
16     id INT,  
17     track_name STRING,  
18     size_bytes INT,  
19     currency STRING,  
20     price FLOAT,  
21     rating_count_tot INT,  
22     rating_count_ver INT,  
23     user_rating FLOAT,  
24     user_rating_ver FLOAT,  
25     ver STRING,  
26     cont_rating STRING,  
27     prime_genre STRING,  
28     sup_devices_num INT,  
29     ipadSc_urls_num INT,  
30     lang_num INT,  
31     vpp_lic INT  
32 );  
33
```

↩ Results

~ Chart

status	
1	Table APP_STORE_DATA successfully created.

- Create a file format named my_csv_format for loading CSV files

```
37
38 CREATE OR REPLACE FILE FORMAT my_csv_format
39 TYPE = 'CSV'
40 FIELD_DELIMITER = ','
41 SKIP_HEADER = 1
42 NULL_IF = ('', 'NULL')
43 FIELD_OPTIONALLY_ENCLOSED_BY = ''
44 DATE_FORMAT = 'AUTO'
45 VALIDATE_UTF8 = TRUE
46 ERROR_ON_COLUMN_COUNT_MISMATCH = TRUE;
47
48
```

↩ Results ~ Chart

	status
1	File format MY_CSV_FORMAT successfully created.

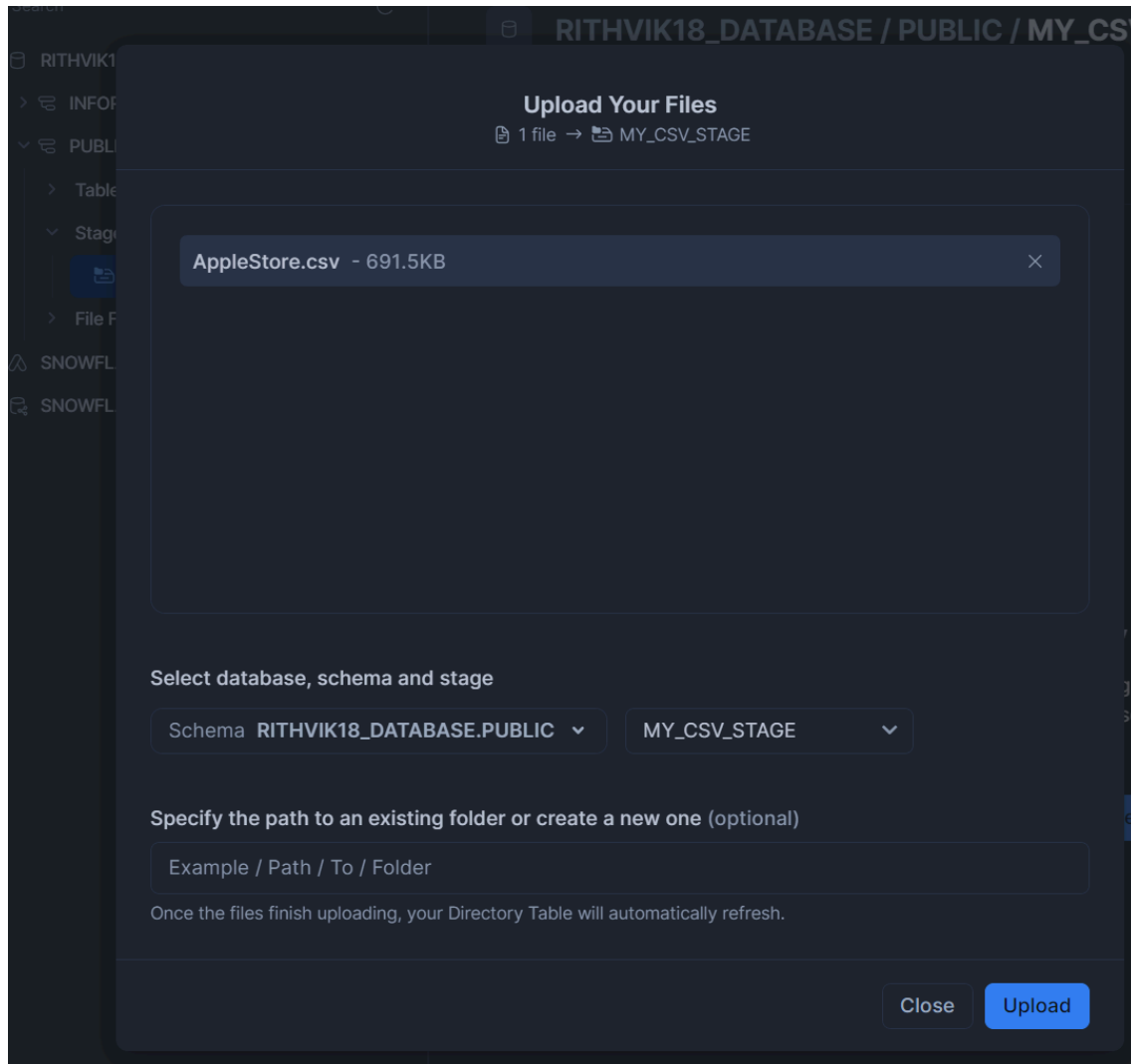
- Create a stage named my_csv_Stage for storing data.
- This stage will act as an intermediate storage for loading or unloading data files

```
48
49
50 CREATE OR REPLACE STAGE my_csv_stage
51 FILE_FORMAT = my_csv_format;
52
53
```

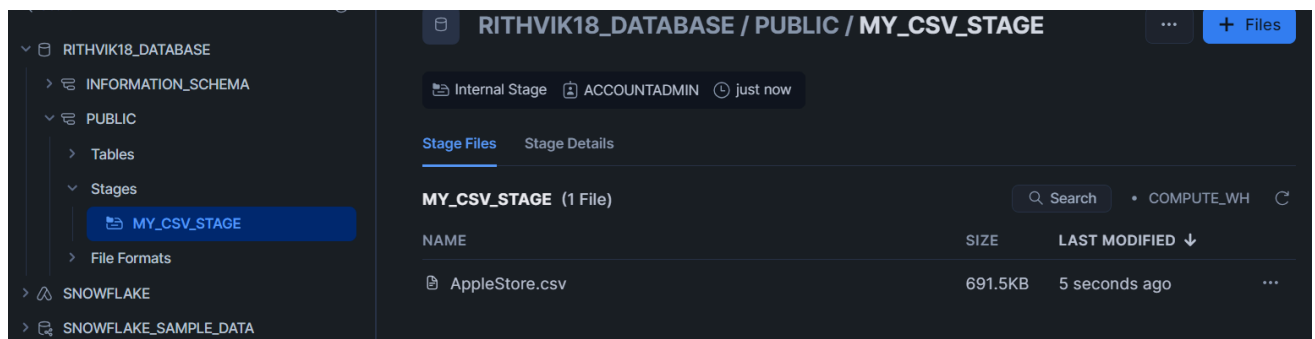
↩ Results ~ Chart

	status
1	Stage area MY_CSV_STAGE successfully created.

- Uploading data into stage using Snowflake UI



- Data Uploaded Successfully into the stage



- The 'List' command is used to list all files in the stage

```
55 | list @my_csv_stage;
```

	name	size	md5	last_modified
1	my_csv_stage/AppStore.csv	708112	56e6ead21417ab7f63ef1a56f7a40754	Sun, 22 Sep 2024 12:04:55

- The Copy command is used to load data from the stage into a table

```
59 | COPY INTO APP_STORE_DATA
60 | FROM @my_csv_stage
61 | FILE_FORMAT = (FORMAT_NAME = 'my_csv_format');
```

	file	status	rows_parsed	rows_loaded	error_limit	errors_seen	first_error1
1	my_csv_stage/AppStore.csv	LOADED	7197	7197	1	0	null

- The data loaded successfully

```
68 | select * from app_store_data;
```

	ID	TRACK_NAME	SIZE_BYTES	CURRENCY	PRICE	RATING_COUNT_TOT	RATIF
1	284882215	Facebook	389879808	USD	0	2974676	
2	389801252	Instagram	113954816	USD	0	2161558	
3	529479190	Clash of Clans	116476928	USD	0	2130805	
4	420009108	Temple Run	65921024	USD	0	1724546	
5	284035177	Pandora - Music & Radio	130242560	USD	0	1126879	
6	429047995	Pinterest	74778624	USD	0	1061624	
7	282935706	Bible	92774400	USD	0	985920	
8	553834731	Candy Crush Saga	222846976	USD	0	961794	
9	324684580	Spotify Music	132510720	USD	0	878563	
10	343200656	Angry Birds	175966208	USD	0	824451	
11	512939461	Subway Surfers	156038144	USD	0	706110	
12	362949845	Fruit Ninja Classic	104590336	USD	1.99	698516	

Query Details

Query duration 887ms

Rows 72K

Query ID 01b73356-3201-393e-0...

Show more

ID 281656475 1188375727

TRACK_NAME 100% filled

SIZE_BYTES

Data Transformation

- Missing values in the user_rating column were replaced with the average rating across all apps.
- Missing values in the price column were replaced with 0 (for free apps).

```
UPDATE app_store_data
SET user_rating = (SELECT AVG(user_rating) FROM app_store_data)
WHERE user_rating IS NULL;

UPDATE app_store_data
SET price = 0
WHERE price IS NULL;
```

- Removed rows where critical columns like track_name, price, or prime_genre were missing to ensure data consistency.

```
81
82 DELETE FROM app_store_data
83 WHERE track_name IS NULL
84 OR prime_genre IS NULL;
85
86
```

- A new column price_category was created to classify apps as either 'Free' or 'Paid' based on their price.

```
86
87 ALTER TABLE app_store_data ADD COLUMN price_category VARCHAR(10);
88
89 UPDATE app_store_data
90 SET price_category = CASE
91     WHEN price = 0 THEN 'Free'
92     ELSE 'Paid'
93 END;
94
95
96
97
```

Results Chart

	number of rows updated	number of multi-joined rows updated
1	7197	0

Query Details

Query duration 857ms

Rows 1

```
67
68
69 select * from app_store_data;
70
71
```

Results Chart

	PRIME_GENRE	SUP_DEVICES_NUM	IPADSC_URLS_NUM	LANG_NUM	VPP_LIC	PRICE_CATEGORY
6	Social Networking	37	5	27	1	Free
7	Reference	37	5	45	1	Free
8	Games	43	5	24	1	Free
9	Music	37	5	18	1	Free
10	Games	38	0	10	1	Free
11	Games	38	5	1	1	Free
12	Games	38	5	13	1	Paid
13	Games	38	4	11	1	Free
14	Games	37	5	10	1	Free
15	Games	38	5	13	1	Free
16	Games	39	5	13	1	Free
17	Games	38	5	18	1	Free

Query Details

Query duration 496ms

Rows 7.2K

Query ID 01b73364-3201-39f3-0...

Show more

ID

281656475 1188375727

TRACK_NAME

100% filled

Insights from Data Analysis

- Average app price: \$1.7 (approx.)
- The average app price (1.7262) is relatively low, suggesting that many apps are either free or priced affordably.

```
96 SELECT AVG(price) AS avg_price,  
97        SUM(rating_count_tot) AS total_ratings,  
98        COUNT(id) AS total_apps  
99 FROM app_store_data;  
100  
101
```

Results Chart

	AVG_PRICE	TOTAL_RATINGS	TOTAL_APPS
1	1.726217869	92790253	7197

- Total number of user ratings: 92.79
- The total number of ratings (92.79 million) suggests a high level of user engagement and activity on the app store.
- Average user rating: 3.5 out of 5
- The average user rating of 3.5 out of 5 indicates a generally positive user experience

```
102 SELECT AVG(user_rating) AS avg_user_rating,  
103        SUM(rating_count_tot) AS total_ratings  
104 FROM app_store_data;  
105  
106  
107
```

Results Chart

AVG_USER_RATING	TOTAL_RATINGS
3.526955676	92790253

- The most popular app genre based on total ratings is Games, with 52,878,491 ratings. This suggests that games are highly rated and popular among users on the app store.

```

105
106 SELECT prime_genre, SUM(rating_count_tot) AS total_ratings
107 FROM app_store_data
108 GROUP BY prime_genre
109 ORDER BY total_ratings DESC
110 LIMIT 1;
111

```

Results Chart

	PRIME_GENRE	TOTAL_RATINGS
1	Games	52878491

- The highest-rated track is "Head Soccer" with a user rating of 5. This suggests that "Head Soccer" is the most popular and well-liked track among users based on the available data.

```

112
113 SELECT track_name, user_rating
114 FROM app_store_data
115 ORDER BY user_rating DESC
116 LIMIT 1;
117

```

Results Chart

	TRACK_NAME	USER_RATING
1	Head Soccer	5

- Paid apps have higher average user ratings than free apps. This could be attributed to factors like perceived quality, reduced ads, and user commitment. However, further analysis is needed to understand the full impact of these factors.

```

118
119 SELECT price_category, AVG(user_rating) AS avg_user_rating
120 FROM app_store_data
121 GROUP BY price_category;
122
123

```

	PRICE_CATEGORY	AVG_USER_RATING
1	Free	3.376725838
2	Paid	3.720948742

- Games is the most popular genre with 3862 apps, followed by Entertainment and Education. This suggests a strong demand for gaming and entertainment apps on the app store.

```


123
124 SELECT prime_genre, COUNT(id) AS total_apps
125 FROM app_store_data
126 GROUP BY prime_genre
127 ORDER BY total_apps DESC;
128
129

```

	PRIME_GENRE	TOTAL_APPS
1	Games	3862
2	Entertainment	535
3	Education	453
4	Photo & Video	349
5	Utilities	248
6	Health & Fitness	180
7	Productivity	178
8	Social Networking	167
9	Lifestyle	144
10	Music	138
11	Shopping	122
12	Sports	114

- The highest-rated prime genre is Productivity with an average user rating of 4.0056. This suggests that productivity apps are generally well-received by users.

```
135
136 SELECT prime_genre, AVG(user_rating) AS avg_user_rating
137 FROM app_store_data
138 GROUP BY prime_genre
139 ORDER BY avg_user_rating DESC
140 LIMIT 1;
141
142
```

 Results  Chart

	PRIME_GENRE	AVG_USER_RATING
1	Productivity	4.005617978