SWIGGY DATA ANAYSIS

- **Download Data:** Search for and download a Swiggy dataset from Kaggle CSV/Excel format for analysis.

- **Data Format:** The dataset was imported in text format, Ensuring compatibility and ease of manipulation within the mysql environment

- **Clean in SQL:** Import the dataset into a SQL database, handle missing or inconsistent values, and standardize data formats like text fields.

- **Analyze in Power BI:** Import the cleaned data into Power BI to create interactive dashboards with charts and KPIs, showcasing order trends, customer behavior, and popular restaurants or dishes.

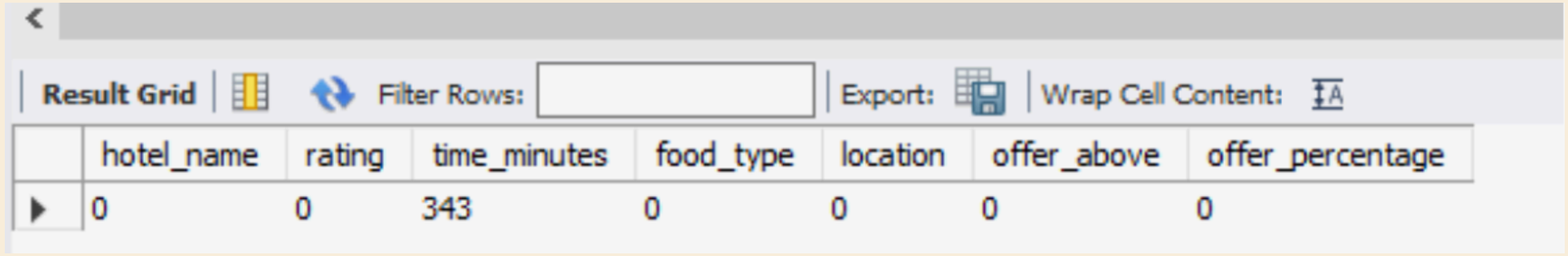
SELECT * FROM SWIGGY;

hotel_name	rating	time_minutes	food_type	location	offer_above	offer_percentage
McDonald's	4.5	27	Burgers, Beverages, Cafe, Desserts	Kandivali East	75	30
KFC	4.2	30	Burgers, Biryani, American, Snacks, Fast Food	Kandivali East	80	40
Domino's Pizza	4.3	25	Pizzas, Italian, Pastas, Desserts	Thakur Village	299	not_available
Charcoal Eats - Biryani & Beyond	4.3	24	Biryani, Kebabs, Hyderabadi, North Indian	Malad Kan East	100	50
Sandwizzaa	4.6	22	Snacks, Fast Food, Beverages, Jain	Kandivali East	120	60
1441 Pizzeria	3.8	32	Italian, Pizzas, Pastas	Kandivali East	120	60
Hangout Cakes & More	4.4	24	Desserts, Bakery	Borivali West	150	60
Love & Cheesecake	4.5	14	Desserts	Kandivali East	not_available	not_available
Nothing But Chicken	4.5	22	Salads, Biryani, Kebabs, Fast Food, Healthy Food	Kandivali East	50	20
Sweet Bengal	4.6	13	Sweets, Indian, Snacks, Desserts	Kandivali East	50	20
Atta Girl	4.7	27	South Indian	Malad West	not_available	not_available
NIC Ice Creams	4.7	17	Ice Cream, Desserts	Kandivali East	149	not_available
Poetry By Love & Cheesecake	4.4	34	Desserts	Kandivali West	not_available	not_available
Havmor Havfunn Ice Cream	4.6	18	Ice Cream, Ice Cream Cakes, Desserts, Bevera...	Malad Kan East	not_available	not_available
NOTO - Ice Creams & Desserts	4.6	15	Ice Cream	Kandivali East	799	not_available
McDonald's Gourmet Burger C...	4.3	26	Burgers, Beverages, Cafe, Desserts	Malad Kan East	not_available	not_available
Grameen Kulfi	4.7	16	Ice Cream, Desserts	Kandivali East	99	not_available
Surbhi Sweets And Snacks	4.2	24	Sweets, Chaat, Snacks, Desserts, Beverages, I...	Dahisar East	599	not_available

1) CHECKING THE NULL VALUES

Ist Method

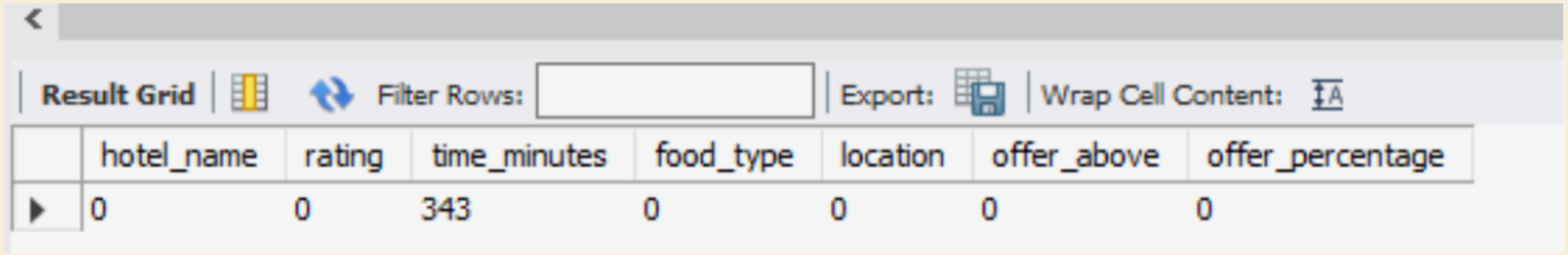
```
SELECT
SUM(CASE WHEN HOTEL_NAME = '' THEN 1 ELSE 0 END ) AS HOTEL_NAME,
SUM(CASE WHEN RATING = '' THEN 1 ELSE 0 END ) AS RATING,
SUM(CASE WHEN TIME_MINUTES = '' THEN 1 ELSE 0 END ) AS TIME_MINUTES,
SUM(CASE WHEN FOOD_TYPE = '' THEN 1 ELSE 0 END ) AS FOOD_TYPE,
SUM(CASE WHEN LOCATION = '' THEN 1 ELSE 0 END ) AS LOCATION,
SUM(CASE WHEN OFFER_ABOVE = '' THEN 1 ELSE 0 END ) AS OFFER_ABOVE,
SUM(CASE WHEN OFFER_PERCENTAGE = '' THEN 1 ELSE 0 END ) AS
OFFER_PERCENTAGE
FROM SWIGGY;
```



hotel_name	rating	time_minutes	food_type	location	offer_above	offer_percentage
0	0	343	0	0	0	0

TO EFFICIENTLY CHECK FOR NULL VALUES IN A DATASET WITH MANY COLUMNS, USE CONCAT() AND GROUP_CONCAT()

```
delimiter //
create procedure  count_blank_rows()
begin
 select group_concat(
  concat('sum(case when`', column_name, '`="" Then 1 else 0 end) as `', column_name ,'`')
  ) into @sql
  from information_schema.columns  where table_name= 'swiggy';
 set @sql = concat('select ', @sql,' from swiggy');

 prepare smt from  @sql;
 execute  smt ;
 deallocate  prepare smt;
end
//
call count_blank_rows() ;
```



hotel_name	rating	time_minutes	food_type	location	offer_above	offer_percentage
0	0	343	0	0	0	0

2. DATA CORRECTION FOR TIME_MINUTES COLUMN

Upon reviewing the dataset, it was found that the values in the "Rating" column represent minutes. To correct this, we will move these values to the "TIME_MINUTES" column, which is currently empty.

```
select hotel_name , rating
, time_minutes from
swiggy where rating like
'%mins%';
```

Result Grid | Filter Rows:

hotel_name	rating	time_minutes
Dominic Pizza	33 mins	
Tiffin Box	27 mins	
Burger Hub	34 mins	
Ghar Ki Rasoi	32 mins	
Iceberg Organic Icecreams	36 mins	
Shahi Darbar	41 mins	
Foodaholic	53 mins	
Cake Works	46 mins	
THE CAKE POINT	34 mins	
Times Of Paratha	40 mins	

THE "RATING" COLUMN CONTAINS VALUES LIKE "33MINS." WE WILL CREATE A FUNCTION TO EXTRACT THE NUMERIC PART (E.G., "33") AND STORE IT IN THE "TIME_MINUTES" COLUMN.

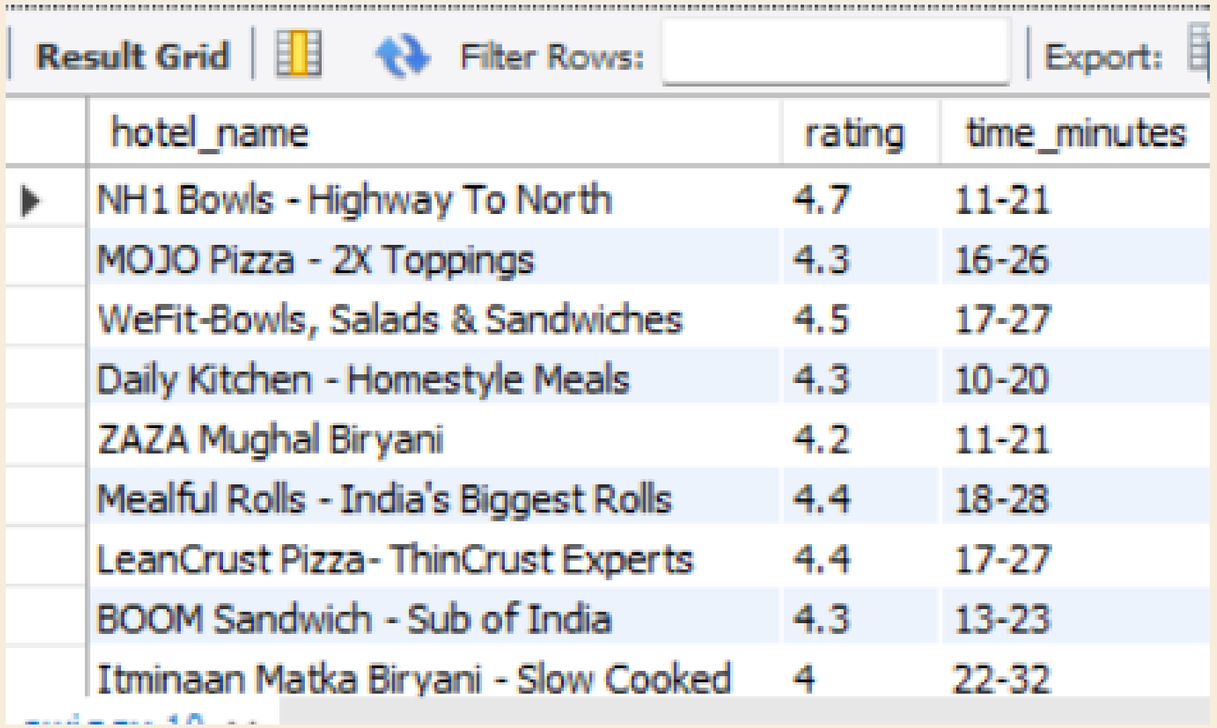
```
delimiter $$
create function f_name(input_str varchar(255))
returns int
deterministic
begin
    return cast(substring_index(input_str, '-', 1) as
unsigned);
end $$
delimiter ;
```

```
create table cleaned as
select * , f_name(rating) as f1 from clean ;

update swiggy as s
inner join cleaned as c
on s.hotel_name = c.hotel_name
set s.time_minutes= c.f1;

select * from swiggy where rating like '%mins%';
```

hotel_name	rating	time_minutes	food_type	location	offer_above	offer_percentage
Dominic Pizza	33 mins	33	Fast Food, Pizzas, Chinese, Italian, Pastas, Des...	Pizza Story	100	50
Tiffin Box	27 mins	27	Biryani, Beverages, North Eastern	Malad Kan East	75	30
Burger Hub	34 mins	34	Burgers, Snacks, Chinese	Borivali	not_available	not_available
Ghar Ki Rasoi	32 mins	32	North Indian	Malad Kan East	not_available	not_available
Iceberg Organic Icecreams	36 mins	36	Ice Cream	Borivali	299	not_available
Shahi Darbar	41 mins	41	Beverages, Desserts, Ice Cream	Malad Kan West	not_available	not_available
Foodaholic	53 mins	53	Chinese	Jog Gor East	999	not_available
Cake Works	46 mins	46	Indian	Malad Kan East	150	30
THE CAKE POINT	34 mins	34	Bakery	Borivali	not_available	not_available
Times Of Paratha	40 mins	40	Indian, Snacks	Kandivali East	100	50
Makhani Darbar	23 mins	23	North Indian, Kebabs, Mughlai, Beverages, Des...	Thakur Village	100	50

During dataset examination, it was found that the TIME_MINUTES column contains range values like '15-20'. To address this, we will calculate and use the average value (e.g., 17.5) for consistency.

```
select hotel_name , rating
, time_minutes  from
swiggy where
time_minutes like '%-%';
```



Result Grid | Filter Rows: | Export:

hotel_name	rating	time_minutes
NH1 Bowls - Highway To North	4.7	11-21
MOJO Pizza - 2X Toppings	4.3	16-26
WeFit-Bowls, Salads & Sandwiches	4.5	17-27
Daily Kitchen - Homestyle Meals	4.3	10-20
ZAZA Mughal Biryani	4.2	11-21
Mealful Rolls - India's Biggest Rolls	4.4	18-28
LeanCrust Pizza- ThinCrust Experts	4.4	17-27
BOOM Sandwich - Sub of India	4.3	13-23
Itminaan Matka Biryani - Slow Cooked	4	22-32

Since we previously created the f_name() function to extract the front part of values, we will now use the l_name() function to extract the last part from range values like '15-20'.

```
delimiter $$
create function l_name(input_str varchar(255))
returns int
deterministic
begin
    return cast(substring_index(input_str, '-', -1)
as unsigned);
end $$
delimiter ;
```

```
create table cleaned as
select * ,f_name(time_minutes) as
f1,l_name(time_minutes) as f2 from clean
where time_minutes like '%-%';

update swiggy as s
inner join cleaned as c
on s.hotel_name = c.hotel_name
set s.time_minutes= ((c.f1 +c.f2)/2);
select * from swiggy where hotel_name =
'MOJO Pizza - 2X Toppings';
```

hotel_name	rating	time_minutes	food_type	location	offer_above	offer_percentage
MOJO Pizza - 2X Toppings	4.3	21.000000000	Pizzas, Italian, Fast Food, Desserts	Kandivali East	100	50

3. DATA CORRECTION FOR RATING COLUMN

UPON EXAMINING THE DATASET, IT WAS DISCOVERED THAT THE RATING COLUMN CONTAINS VALUES REPRESENTING ‘MINS’ . TO RECTIFY THIS, WE WILL TAKE THE AVERAGE WITH LOCATION

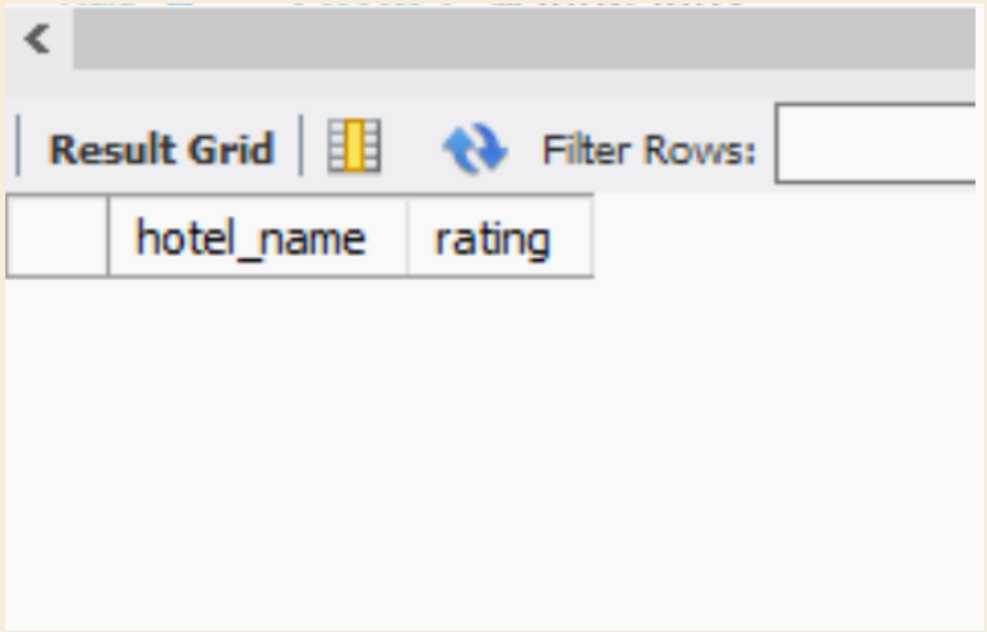
```
select * from swiggy
where rating like
'%mins%';
```

hotel_name	rating	time_minutes
Dominic Pizza	33 mins	33
Tiffin Box	27 mins	27
Burger Hub	34 mins	34
Ghar Ki Rasoi	32 mins	32
Iceberg Organic Icecreams	36 mins	36
Shahi Darbar	41 mins	41
Foodaholic	53 mins	53

```
update swiggy as s
inner join (
select location, round(avg(rating),2) as
average
from swiggy
where rating not like '%mins%'
group by location ) as t
on s.location = t.location
set s.rating = t.average
where s.rating like '%mins%';
```

FOR LOCATIONS THAT APPEAR ONLY ONCE, RATINGS WILL BE FILLED WITH THE AVERAGE VALUE OF THE "RATING" COLUMN .

```
set @tot_Avg = (select round(avg(rating),2)
as average
from swiggy
where rating not like '%mins%'
);
update  swiggy
set rating = @tot_avg
where rating like '%mins%';

select hotel_name,rating
from swiggy where
rating like '%mins%' ;
```



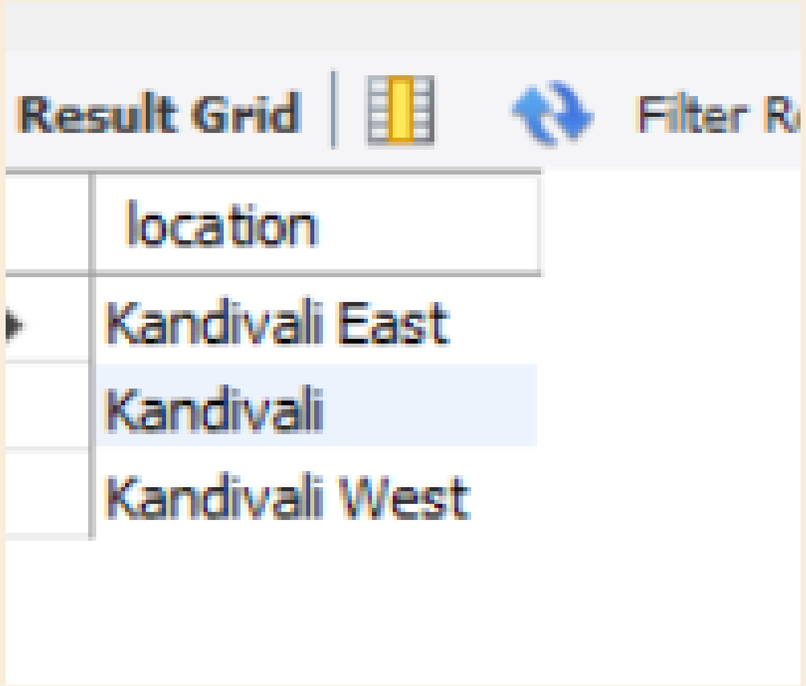
4. DATA CORRECTION FOR LOCATION COLUMN

Locations such as "Kandivali East" , "Kandivali (East),Thakur Village,"Thakur Village, Kandivali (E)" all represent "Kandivali East". Similarly, variations for "Kandivali West" also

```
select distinct(location) from swiggy
 where  location like '%Kandivali%' ;
```

location
Kandivali East
Kandivali West
Thakur village Kandivali east
Thakur Village, Kandivali (E)
Kandivali
Mahavir Nagar Kandivali West
Kandivali (East), Thakur Village
Kandivali East, Thakur Village
Kandivali East Lokhandwala
Kandivali borivali East
Kandivali, Malad East
Kandivali - West

```
update swiggy
set location  ='Kandivali East'
where location like '%East%' and location like '%Kandivali%';

update swiggy
set location  ='Kandivali West'
where location like '%West%' and location like '%Kandivali%';

update swiggy
set location  ='Kandivali East'
where location like '%E%' and location like '%Kandivali%';

update swiggy
set location  ='Kandivali West'
where location like '%W%' and location like '%Kandivali%';

select distinct(location) from swiggy where location like '%Kandivali%' ;
```



5. DATA CORRECTION FOR OFFER_PERCENTAGE COLUMN

IF THE "OFFER_ABOVE" VALUE IS NOT AVAILABLE , THE "OFFER_PERCENTAGE" WILL BE SET TO 0 TO MAINTAIN DATA CONSISTENCY.

```
select offer_above,offer_percentage
from swiggy ;
```

offer_above	offer_percentage
not_available	not_available
50	20
50	20
not_available	not_available
149	not_available
not_available	not_available
not_available	not_available
799	not_available

```
update swiggy
set offer_percentage = 0
where  offer_above =
'not_available' ;
```

offer_above	offer_percentage
100	50
50	20
not_available	0
500	10
700	10
not_available	0
not_available	0
40	10
not_available	0
75	30

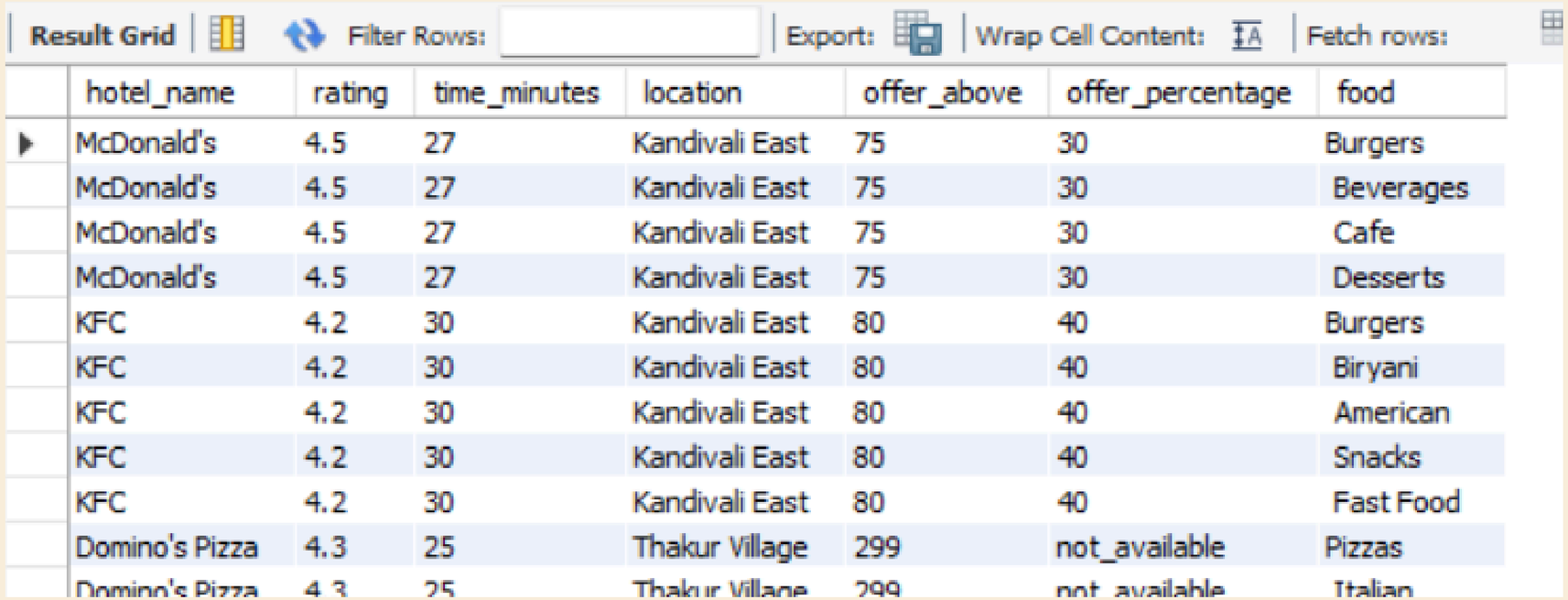
6. DATA CORRECTION FOR FOOD_TYPE COLUMN

```
select food_type from
swiggy;
```

food_type
Burgers, Beverages, Cafe, Desserts
Burgers, Biryani, American, Snacks, Fast Food
Pizzas, Italian, Pastas, Desserts
Biryani, Kebabs, Hyderabadi, North Indian
Snacks, Fast Food, Beverages, Jain
Italian, Pizzas, Pastas
Desserts, Bakery
Desserts
Salads, Biryani, Kebabs, Fast Food, Healthy Food
Sweets, Indian, Snacks, Desserts
South Indian

```
select hotel_name ,rating,time_minutes,location,offer_above,offer_percentage,
substring_index( substring_index(food_type ,',',numbers.n),',', -1) as 'food'
from  swiggy
 join
 (
  select 1+a.N + b.N*10 as n from
  (
   (
   SELECT 0 AS N UNION ALL SELECT 1 UNION ALL SELECT 2 UNION ALL SELECT 3 UNION
ALL SELECT 4 UNION ALL SELECT 5 UNION ALL SELECT 6 UNION ALL SELECT 7
   UNION ALL SELECT 8 UNION ALL SELECT 9) a
   cross join
   (
   SELECT 0 AS N UNION ALL SELECT 1 UNION ALL SELECT 2 UNION ALL SELECT 3 UNION
ALL SELECT 4 UNION ALL SELECT 5 UNION ALL SELECT 6 UNION ALL SELECT 7
   UNION ALL SELECT 8 UNION ALL SELECT 9)b
  )
 )  as numbers
    on  char_length(food_type)  - char_length(replace(food_type ,',','')) >= numbers.n-1
    )
```

OUTPUT :



Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

hotel_name	rating	time_minutes	location	offer_above	offer_percentage	food
McDonald's	4.5	27	Kandivali East	75	30	Burgers
McDonald's	4.5	27	Kandivali East	75	30	Beverages
McDonald's	4.5	27	Kandivali East	75	30	Cafe
McDonald's	4.5	27	Kandivali East	75	30	Desserts
KFC	4.2	30	Kandivali East	80	40	Burgers
KFC	4.2	30	Kandivali East	80	40	Biryani
KFC	4.2	30	Kandivali East	80	40	American
KFC	4.2	30	Kandivali East	80	40	Snacks
KFC	4.2	30	Kandivali East	80	40	Fast Food
Domino's Pizza	4.3	25	Thakur Village	299	not_available	Pizzas
Domino's Pizza	4.3	25	Thakur Village	299	not_available	Italian

THE SWIGGY_CLEANED TABLE SERVES AS THE FINALIZED, CLEANED DATASET USED FOR ANALYSIS IN POWER BI

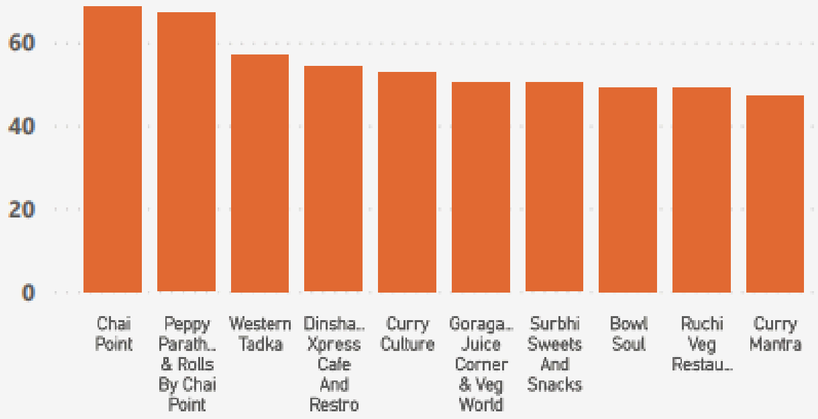
SWIGGY ANALYSIS DASHBOARD

hotel_name: All

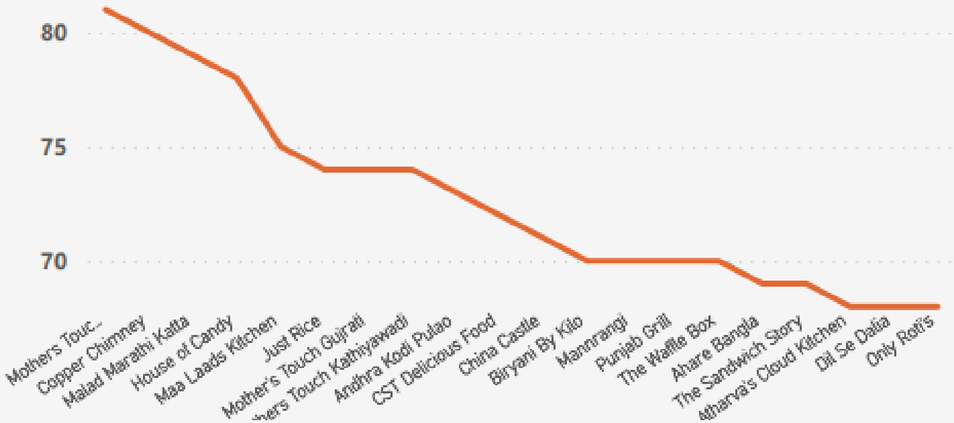
location: All

Total Hotels	Average Rating	Avg Delivery Time	Food Items
1729	4	42	142

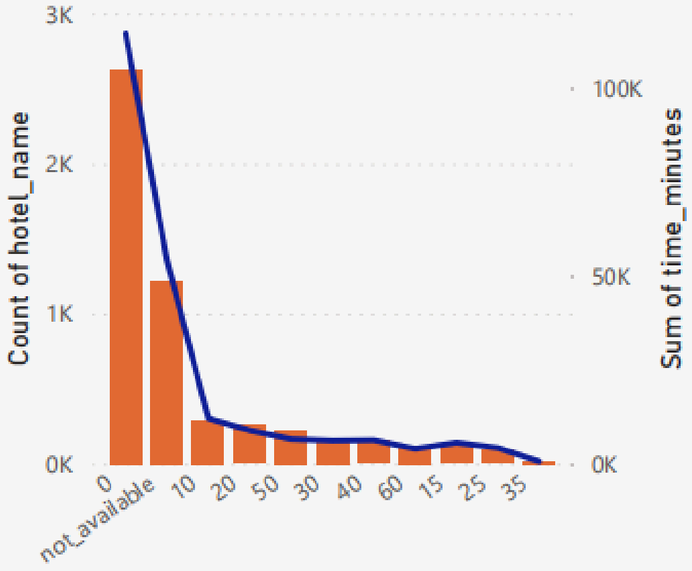
Rating by hotel



Average of time_minutes by hotel

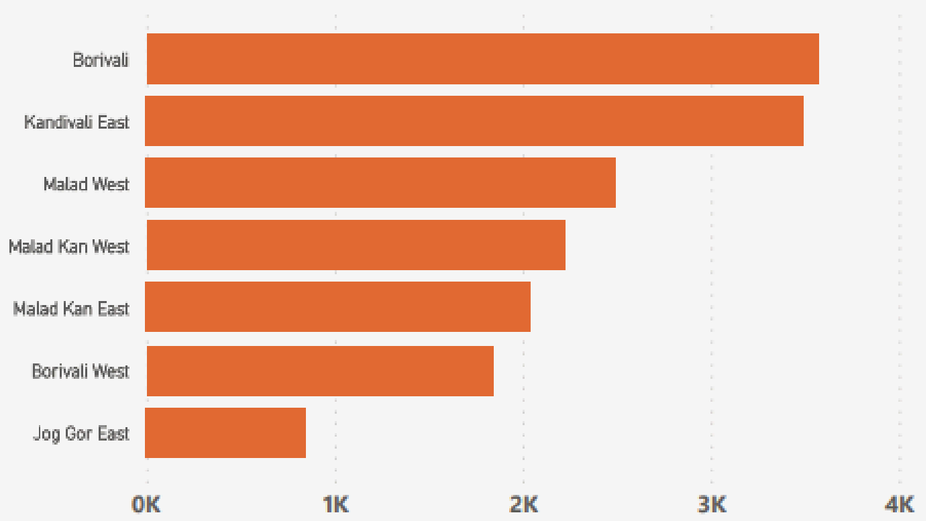


Hotels and Duration by offer_percentage



Number of hotels by location

Rating by location



Rating by food

INSIGHTS

Top Hotels by Rating:

- Hotels such as Chai Point, Peppy Paratha, and Western Tadka have the highest ratings.

Average Delivery Time by Hotel:

- Delivery time varies significantly across hotels, with some averaging around 80 minutes and others below 70 minutes.

Offer Percentage Impact:

- The majority of hotels fall under the "not-available" category for offers.
- Hotels with offer percentages of 10-20% have a significant presence, while higher offer percentages are less common.

Geographical Distribution:

- A high concentration of hotels is observed in Mumbai, Bengaluru, and Hyderabad.

Top Locations by Ratings:

- Borivali and Kandivali East are the top-rated locations, followed by Malad West and Malad Kan West.

Food Category Ratings:

- Beverages, Snacks, and Indian Cuisine dominate in terms of customer ratings.
- Subcategories like Desserts, Bakery, and Chinese also perform well.

THANK YOU!