**GROUP PROJECT: QUERY LOG**

**MARKET ANALYSIS: Query to get the CBGs with highest youth ratio (>= 80%) among the top 5 counties we selected:**

SELECT

CountyName,

cbg,

YouthRatio

from

(

SELECT

\*,

ROW\_NUMBER() OVER (

PARTITION BY CountyName

ORDER BY YouthRatio DESC

) rn

FROM

(

select

case when cbg like '18157%' then 'Tippecanoe County' when cbg like '17019%' then 'Champaign County' when cbg like '42027%' then 'Centre County' when cbg like '18105%' then 'Monroe County' when cbg like '28071%' then 'Lafayette County' end as CountyName,

cbg,

(

sum(`pop\_m\_18-19`)+ sum(`pop\_f\_18-19`)+ sum(`pop\_m\_20`)+ sum(`pop\_f\_20`)+ sum(`pop\_m\_21`)+ sum(`pop\_f\_21`)+ sum(`pop\_m\_22-24`)+ sum(`pop\_f\_22-24`)

)/ sum(pop\_total) as YouthRatio

FROM

`antilles-data-mgmt58200-final.safegraph.cbg\_demographics` d

join `antilles-data-mgmt58200-final.safegraph.cbg\_fips` c on LEFT(d.cbg, 2) = c.state\_fips

and SUBSTRING(d.cbg, 3, 3) = c.county\_fips

WHERE

(

cbg LIKE '18157%'

or cbg LIKE '17019%'

or cbg LIKE '42027%'

or cbg LIKE '18105%'

or cbg LIKE '28071%'

)

and cbg is not null

group by

cbg

having

youthRatio >= 0.90

order by

youthratio desc

)

)

WHERE

rn = 1;

**MARKET ANALYSIS - Query to get income band information for the cbgs with high youth ratio in the specified counties**

select

case when cbg like '18157%' then 'Tippecanoe County' when cbg like '17019%' then 'Champaign County' when cbg like '42027%' then 'Centre County' when cbg like '18105%' then 'Monroe County' when cbg like '28071%' then 'Lafayette County' end as CountyName,

d.cbg,

sum(`inc\_lt10`) Lt10,

sum(`inc\_10-15`) Inc10To15,

sum(`inc\_15-20`) Inc15To20,

sum(`inc\_20-25`) Inc20To25,

sum(`inc\_25-30`) Inc25To30,

sum(`inc\_30-35`) Inc30To35,

sum(`inc\_35-40`) Inc35To40,

sum(`inc\_40-45`) Inc40To45,

sum(`inc\_45-50`) Inc45To50,

sum(`inc\_50-60`) Inc50To60,

sum(`inc\_60-75`) Inc60To75,

sum(`inc\_75-100`) Inc75To100,

sum(`inc\_100-125`) Inc100To125,

sum(`inc\_125-150`) Inc125To150,

sum(`inc\_150-200`) Inc150To200

FROM

`antilles-data-mgmt58200-final.safegraph.cbg\_demographics` d

join `antilles-data-mgmt58200-final.safegraph.cbg\_fips` c on LEFT(d.cbg, 2) = c.state\_fips

and SUBSTRING(d.cbg, 3, 3) = c.county\_fips

WHERE

cbg in (

select

cbg

from

(

select

cbg,

(

sum(`pop\_m\_18-19`)+ sum(`pop\_f\_18-19`)+ sum(`pop\_m\_20`)+ sum(`pop\_f\_20`)+ sum(`pop\_m\_21`)+ sum(`pop\_f\_21`)+ sum(`pop\_m\_22-24`)+ sum(`pop\_f\_22-24`)

)/ sum(pop\_total) as YouthRatio

FROM

`antilles-data-mgmt58200-final.safegraph.cbg\_demographics` d

join `antilles-data-mgmt58200-final.safegraph.cbg\_fips` c on LEFT(d.cbg, 2) = c.state\_fips

and SUBSTRING(d.cbg, 3, 3) = c.county\_fips

WHERE

(

cbg LIKE '18157%'

or cbg LIKE '17019%'

or cbg LIKE '42027%'

or cbg LIKE '18105%'

or cbg LIKE '28071%'

)

group by

cbg

having

youthRatio >= 0.80

order by

youthratio desc

)

)

group by cbg

**LOCATION ANALYSIS – Query to retrieve the bottom 10 restaurants by footfall, along with their details, located in areas with high youth density. These restaurants should also be in counties where the overall footfall is above the county average. Its done similarly for each county.**

WITH RestaurantCounts AS (

SELECT

'Tippecanoe County' AS county,

cd.cbg,

COUNT(p.safegraph\_place\_id) AS restaurant\_count

FROM

`antilles-data-mgmt58200-final.safegraph.cbg\_demographics` cd

JOIN

`antilles-data-mgmt58200-final.safegraph.visits` v ON v.poi\_cbg = cd.cbg

JOIN

`antilles-data-mgmt58200-final.safegraph.places` p ON p.safegraph\_place\_id = v.safegraph\_place\_id

WHERE

p.top\_category LIKE '%Restaurant%'

AND cd.cbg LIKE '18157%'

GROUP BY

cd.cbg

),

PopulationData AS (

SELECT

'Tippecanoe County' AS county,

cd.cbg,

SUM(cd.pop\_total) AS population,

(SUM(cd.`pop\_m\_18-19`) + SUM(cd.`pop\_f\_18-19`) +

SUM(cd.`pop\_m\_20`) + SUM(cd.`pop\_f\_20`) +

SUM(cd.`pop\_m\_21`) + SUM(cd.`pop\_f\_21`) +

SUM(cd.`pop\_m\_22-24`) + SUM(cd.`pop\_f\_22-24`)) / SUM(cd.pop\_total) AS youth\_ratio

FROM

`antilles-data-mgmt58200-final.safegraph.cbg\_demographics` cd

WHERE

cd.cbg LIKE '18157%'

GROUP BY

cd.cbg

),

FootfallData AS (

SELECT

v.poi\_cbg AS cbg,

'Tippecanoe County' AS county,

SUM(v.raw\_visit\_counts) AS total\_footfall

FROM

`antilles-data-mgmt58200-final.safegraph.visits` v

WHERE

v.poi\_cbg LIKE '18157%'

GROUP BY

v.poi\_cbg

),

AverageFootfall AS (

SELECT

'Tippecanoe County' AS county,

AVG(total\_footfall) AS avg\_footfall

FROM

FootfallData

),

DensityData AS (

SELECT

rc.county,

rc.cbg,

rc.restaurant\_count,

pd.population,

pd.youth\_ratio,

fd.total\_footfall,

af.avg\_footfall,

(rc.restaurant\_count / NULLIF(pd.population, 0)) AS restaurant\_density,

ROW\_NUMBER() OVER (ORDER BY (rc.restaurant\_count / NULLIF(pd.population, 0)) ASC) AS density\_rank

FROM

RestaurantCounts rc

LEFT JOIN

PopulationData pd ON rc.cbg = pd.cbg

LEFT JOIN

FootfallData fd ON rc.cbg = fd.cbg

LEFT JOIN

AverageFootfall af ON rc.county = af.county

WHERE

rc.restaurant\_count > 0

AND pd.youth\_ratio >= 0.80

AND fd.total\_footfall > af.avg\_footfall

)

SELECT

d.county,

v.location\_name,

v.raw\_visit\_counts AS footfall,

p.naics\_code,

v.poi\_cbg AS cbg,

d.avg\_footfall -- Include the average footfall for each area

FROM

`antilles-data-mgmt58200-final.safegraph.visits` v

JOIN

`antilles-data-mgmt58200-final.safegraph.places` p ON v.safegraph\_place\_id = p.safegraph\_place\_id

JOIN

DensityData d ON v.poi\_cbg = d.cbg

WHERE

p.top\_category LIKE '%Restaurant%'

ORDER BY

footfall ASC

LIMIT 10;

**BRAND APPEAL: Query to identify the popular restaurants (highest footfall)**

WITH RankedRestaurants AS (

SELECT

CASE

WHEN cd.cbg LIKE '18157%' THEN 'Tippecanoe County'

WHEN cd.cbg LIKE '17019%' THEN 'Champaign County'

WHEN cd.cbg LIKE '42027%' THEN 'Centre County'

WHEN cd.cbg LIKE '18105%' THEN 'Monroe County'

WHEN cd.cbg LIKE '28071%' THEN 'Lafayette County'

END AS county,

v.location\_name,

SUM(v.raw\_visit\_counts) AS total\_visits,

ROW\_NUMBER() OVER (PARTITION BY

CASE

WHEN cd.cbg LIKE '18157%' THEN 'Tippecanoe County'

WHEN cd.cbg LIKE '17019%' THEN 'Champaign County'

WHEN cd.cbg LIKE '42027%' THEN 'Centre County'

WHEN cd.cbg LIKE '18105%' THEN 'Monroe County'

WHEN cd.cbg LIKE '28071%' THEN 'Lafayette County'

END

ORDER BY SUM(v.raw\_visit\_counts) DESC

) AS rank

FROM `antilles-data-mgmt58200-final.safegraph.cbg\_demographics` cd

JOIN `antilles-data-mgmt58200-final.safegraph.visits` v

ON v.poi\_cbg = cd.cbg

JOIN

`antilles-data-mgmt58200-final.safegraph.places` p

ON p.safegraph\_place\_id = v.safegraph\_place\_id

WHERE

(cd.cbg LIKE '18157%' OR cd.cbg LIKE '17019%' OR cd.cbg LIKE '42027%' OR

cd.cbg LIKE '18105%' OR cd.cbg LIKE '28071%')

AND p.top\_category LIKE '%Restaurant%'

AND p.location\_name NOT LIKE '%Commons%'

AND EXTRACT(YEAR FROM v.date\_range\_start) = 2020

AND EXTRACT(MONTH FROM v.date\_range\_start) = 1

AND EXTRACT(YEAR FROM v.date\_range\_end) = 2020

AND EXTRACT(MONTH FROM v.date\_range\_end) = 2

GROUP BY

cd.cbg, v.postal\_code, v.location\_name)

SELECT county, location\_name, SUM(total\_visits) AS total\_visits

FROM RankedRestaurants

WHERE rank <= 20

GROUP BY county, location\_name

ORDER BY

county, total\_visits DESC;

**DEMAND ANALYSIS: Query To get footfall for restaurants based on hour for selected counties (before March 2020)**

SELECT hour,SUM(visits) AS total\_visits

FROM (

SELECT

[STRUCT(0 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(0)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(1 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(1)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(2 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(2)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(3 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(3)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(4 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(4)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(5 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(5)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(6 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(6)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(7 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(7)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(8 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(8)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(9 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(9)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(10 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(10)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(11 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(11)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(12 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(12)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(13 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(13)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(14 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(14)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(15 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(15)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(16 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(16)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(17 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(17)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(18 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(18)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(19 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(19)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(20 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(20)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(21 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(21)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(22 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(22)], r'[^0-9]', '') AS INT64) AS visits),

STRUCT(23 AS hour, CAST(REGEXP\_REPLACE(SPLIT(popularity\_by\_hour, ',')[OFFSET(23)], r'[^0-9]', '') AS INT64) AS visits)

] AS hours

FROM

`antilles-data-mgmt58200-final.safegraph.visits` AS v

JOIN

`antilles-data-mgmt58200-final.safegraph.cbg\_demographics` AS d

ON

v.poi\_cbg = d.cbg

join `antilles-data-mgmt58200-final.safegraph.brands` b on v.safegraph\_brand\_ids=b.safegraph\_brand\_id

WHERE (

poi\_cbg LIKE '18157%'

or poi\_cbg LIKE '17019%'

or poi\_cbg LIKE '42027%'

or poi\_cbg LIKE '18105%'

or poi\_cbg LIKE '28071%'

)

and b.top\_category like '%Restaurants%'

AND DATE(date\_range\_end) <= '2020-02-29'

), UNNEST(hours) AS hour\_visits\_struct

GROUP BY

hour\_visits\_struct.hour

ORDER BY

total\_visits DESC;

**DEMAND ANALYSIS: Query to get day-wise footfall for the top restaurants we have selected across the 5 counties:**

select distinct location\_name, sum(Monday\_visits) MondayVisits, sum(Tuesday\_visits) TuesdayVisits, sum(Wednesday\_visits) WednesdayVisits, sum(Thursday\_visits) ThursdayVisits, sum(Friday\_visits) FridayVisits, sum(Saturday\_visits) SaturdayVisits, sum(Sunday\_visits) SundayVisits,

from (

SELECT location\_name,

CAST(JSON\_EXTRACT\_SCALAR(popularity\_by\_day, '$.Monday') AS INT64) AS Monday\_visits,

CAST(JSON\_EXTRACT\_SCALAR(popularity\_by\_day, '$.Tuesday') AS INT64) AS Tuesday\_visits,

CAST(JSON\_EXTRACT\_SCALAR(popularity\_by\_day, '$.Wednesday') AS INT64) AS Wednesday\_visits,

CAST(JSON\_EXTRACT\_SCALAR(popularity\_by\_day, '$.Thursday') AS INT64) AS Thursday\_visits,

CAST(JSON\_EXTRACT\_SCALAR(popularity\_by\_day, '$.Friday') AS INT64) AS Friday\_visits,

CAST(JSON\_EXTRACT\_SCALAR(popularity\_by\_day, '$.Saturday') AS INT64) AS Saturday\_visits,

CAST(JSON\_EXTRACT\_SCALAR(popularity\_by\_day, '$.Sunday') AS INT64) AS Sunday\_visits

FROM `antilles-data-mgmt58200-final.safegraph.visits`

WHERE location\_name IN ('McDonald\'s', 'Starbucks', 'Chick-fil-A', 'Olive Garden',

'Cracker Barrel', 'Texas Roadhouse', 'Panera Bread',

'Chili\'s Grill & Bar', 'Steak \'n Shake', 'Wendy\'s',

'IHOP', 'Culver\'s', 'Cheddar\'s Scratch Kitchen')

AND (poi\_cbg LIKE '18157%' or poi\_cbg LIKE '17019%' or poi\_cbg LIKE '42027%' or poi\_cbg LIKE '18105%' or poi\_cbg LIKE '28071%')

AND EXTRACT(MONTH FROM date\_range\_start) = 2

)

group by location\_name