Group 36
April 5th, 2024

Data Analytics, OLAP Queries & BI Dashboard (Deliverable D)

Group 36

Maria Stancu (300243486)

Trinity Vermeire (300129927)

Bradley Wills (300187635)

Table of Contents

Group 36
April 5th, 2024

Part I Write OLAP queries.

Write OLAP queries against Biketheft database to perform data analysis and trends about bike thefts in the City of Toronto neighborhoods of from different perspectives.

- Connect to PostgreSQL Biketheft database.
- Run the script: **Queries_ Deliverable_D.sql**
    - The SQL file contains the queries listed below

Q1. Rollup/drill down

Count bike theft by year

```
select year, count(theft_id)
from date_dimension d, bike_theft_fact btf
where d.date_key=btf.date_key
group by year
order by year;
```

Drill-down year->month

```
select d.year, d.month , count(theft_id)
from date_dimension d, bike_theft_fact btf
where d.date_key=btf.date_key
group by d.year,d.month
order by d.year,d.month ;
```

Rollup Count bike theft by bike type

```
select bike_type, count(theft_id)
from bike_dimension b, bike_theft_fact btf
where b.bike_key=btf.bike_key
group by bike_type;
```

Drill-down bike_type->bike_model

```
select bike_type, bike_model, count(theft_id)
from bike_dimension b, bike_theft_fact btf
where b.bike_key=btf.bike_key
group by bike_type, bike_model;
```

*Q2. S*lice

Count bike theft by neighborhood for year 2023

```
select n.neighbourhood_name, count(theft_id)
from neighbourhood_dimension n,date_dimension d, bike_theft_fact btf
where n.neighbourhood_key=btf.hood_key
and  d.date_key=btf.date_key
and d.year=2023
group by n.neighbourhood_name;
```

Slice Count bike thefts by year for bike_type 'MT'

```
select d.year, count(theft_id)
from bike_dimension b,date_dimension d, bike_theft_fact btf
where b.bike_key=btf.bike_key
and  d.date_key=btf.date_key
and b.bike_type='MT'
group by d.year
order by d.year;
```

Q4 Dice

Count bike thefts by bike_type,neighborhood for year 2023 and bike_type='MT'

```
select bike_type,neighbourhood_name, count(theft_id)
from bike_dimension b,neighbourhood_dimension n,date_dimension d, bike_theft_fact btf
where b.bike_key=btf.bike_key
and n.neighbourhood_key=btf.hood_key
and  d.date_key=btf.date_key
and d.year=2023 and bike_type='MT'
group by bike_type,neighbourhood_name;
```

Count bike thefts by month,neighborhood for year 2023 and in 'Annex' neighborhood

```
select d.month,neighbourhood_name, count(theft_id)
from bike_dimension b,neighbourhood_dimension n,date_dimension d, bike_theft_fact btf
where b.bike_key=btf.bike_key
and n.neighbourhood_key=btf.hood_key
and  d.date_key=btf.date_key
and d.year=2023 and neighbourhood_name = 'Annex'
group by d.month,neighbourhood_name;
```

Q5. Combining OLAP operations

Compare the number of thefts during the month of July in 2023 and 2022 for the bike type MT

```
select n.neighbourhood_name, b.bike_type , count(theft_id)
from bike_dimension b,neighbourhood_dimension n,date_dimension d, bike_theft_fact btf
where b.bike_key=btf.bike_key
and n.neighbourhood_key=btf.hood_key
and  d.date_key=btf.date_key
and d.year in (2023, 2022)
and d.month =7
and b.bike_type='MT'
group by  n.neighbourhood_name, b.bike_type;
```

*Compare the number of thefts during the summer months (June, July, August) for the following years 2023 ,2022 2021, 2020 f*or the following bike type 'MT','RG', 'SC'

```
select d.month,n.neighbourhood_name, b.bike_type , count(theft_id)
from bike_dimension b,neighbourhood_dimension n,date_dimension d, bike_theft_fact btf
where b.bike_key=btf.bike_key
and n.neighbourhood_key=btf.hood_key
and  d.date_key=btf.date_key
and d.year in (2023, 2022, 2021, 2020)
and d.month  in (6,7,8)
and b.bike_type in ('MT','RG', 'SC')
group by d.month, n.neighbourhood_name, b.bike_type;
```

Compare the number of thefts byneighborhood, average robberies in January and July of 2023

```
select n.neighbourhood_name neighborhood,  n.avg_robbery , count(theft_id) NrOfThefts
from  neighbourhood_dimension n, bike_theft_fact bf, date_dimension d
where n.neighbourhood_key=bf.hood_key
and d.date_key=bf.date_key
and year =2023 and month in(71,) and n.avg_robbery > 40 and n.avg_robbery < 200
group by n.neighbourhood_name,  n.avg_robbery
order by n.neighbourhood_name,  n.avg_robbery;
```

Compare the number of thefts by year, neighborhood, income in the 2023,2022,2021,2020 and neighborhood income between 80000 and 150000

```
select d.year,n.neighbourhood_name neighborhood,  n.avg_income , count(theft_id) NrOfThefts
from  neighbourhood_dimension n,date_dimension d, bike_theft_fact bf
where n.neighbourhood_key=bf.hood_key
and d.date_key=bf.date_key
and n.avg_income > 80000 and n.avg_income <150000
and d.year in (2023,2022,2021,2020)
group by d.year,n.neighbourhood_name,  n.avg_income
order by d.year, n.neighbourhood_name,  n.avg_income;
```

Q6. Iceberg

Count bike theft by bike_type,neighborhood, year if theft count greater than 10

```
select b.bike_type,n.neighbourhood_name, d.year, count(theft_id)
from bike_dimension b,neighbourhood_dimension n,date_dimension d, bike_theft_fact btf
where b.bike_key=btf.bike_key
and n.neighbourhood_key=btf.hood_key
and  d.date_key=btf.date_key
group by b.bike_type,n.neighbourhood_name, d.year
having count(theft_id) >50;

select n.neighbourhood_name, d.year, count(theft_id)
from bike_dimension b,neighbourhood_dimension n,date_dimension d, bike_theft_fact btf
where b.bike_key=btf.bike_key
and n.neighbourhood_key=btf.hood_key
and  d.date_key=btf.date_key
group by b.bike_type,n.neighbourhood_name, d.year
```

```
order by  count(theft_id) desc
limit 10;
```

Q7 Windowing queries

Compare bike typet's price with the average price in its bike type and rank :

```
SELECT bike_type, bike_cost,
     RANK() OVER(PARTITION BY bike_type ORDER BY bike_cost ) Rank
FROM bike_dimension
ORDER BY bike_type,  Rank;
```

Q8 Using the Window clause

RANGE BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW

```
select n.neighbourhood_name,  d.month,count(theft_id) OVER W as movcount
from neighbourhood_dimension n, date_dimension d,bike_theft_fact btf
where  n.neighbourhood_key=btf.hood_key
and d.date_key=btf.date_key
WINDOW W AS (PARTITION BY n.neighbourhood_name
ORDER BY  d.month   RANGE BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW)
```

Q9 Grouping Sets()

Grouping Sets() using bike_type and neighborhood_name

```
select bike_type,neighbourhood_name, count(theft_id)
from bike_dimension b,neighbourhood_dimension n, bike_theft_fact btf
where b.bike_key=btf.bike_key
and n.neighbourhood_key=btf.hood_key
group by Grouping sets ((bike_type),(neighbourhood_name),());
```

*Q10. R*ollup() by year, month

```
select d.year, d.month, grouping(d.year, d.month), count(theft_id)
from date_dimension d, bike_theft_fact btf
where d.date_key=btf.date_key
group by ROLLUP(d.year, d.month)
order by d.year,d.month;
```

--rollup() by bike_type, bike_model

```
select bike_type, bike_model, grouping(bike_type, bike_model), count(theft_id)
from bike_dimension b, bike_theft_fact btf
where b.bike_key=btf.bike_key
group by ROLLUP(bike_type, bike_model);
```

*Q11 C*ube

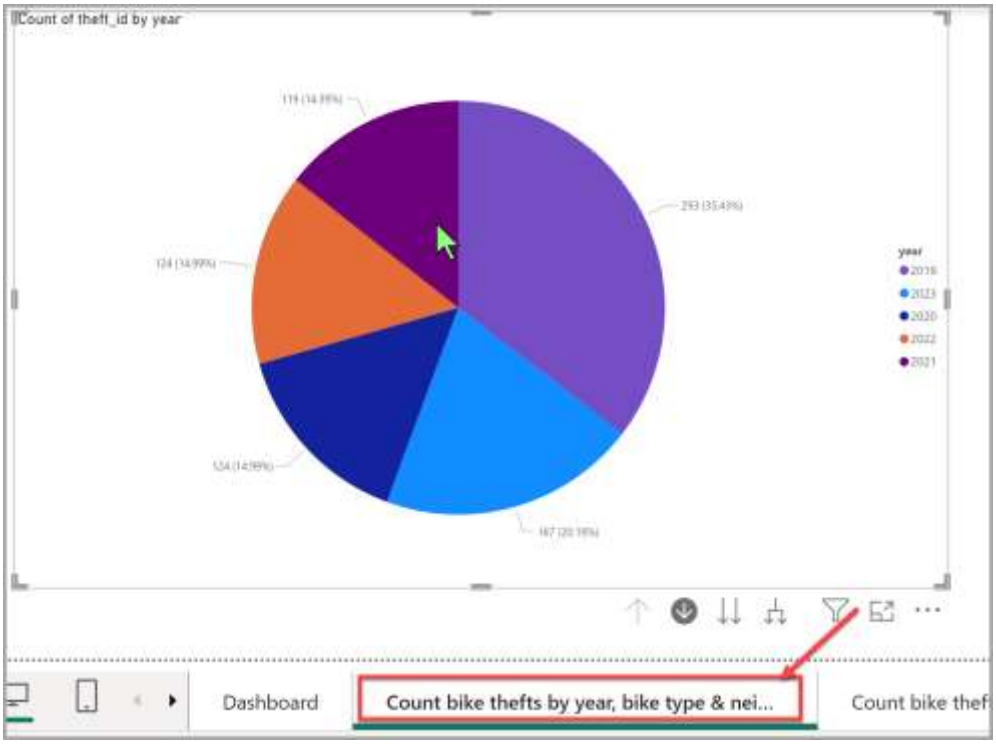
```
select bike_type,neighbourhood_name,d.year, count(theft_id)
from bike_dimension b,neighbourhood_dimension n, date_dimension d,bike_theft_fact btf
where b.bike_key=btf.bike_key
and n.neighbourhood_key=btf.hood_key
```

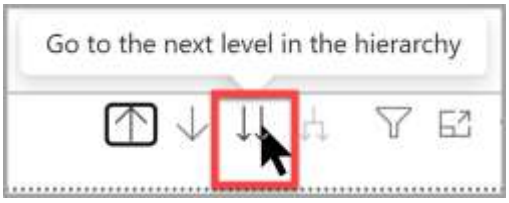
```
and d.date_key=btf.date_key
group by cube(bike_type,neighbourhood_name,d.year)


select d.year, neighbourhood_name,bike_type,count(theft_id)
from bike_dimension b,neighbourhood_dimension n, date_dimension d,bike_theft_fact btf
where b.bike_key=btf.bike_key
and n.neighbourhood_key=btf.hood_key
and d.date_key=btf.date_key
group by cube(d.year,neighbourhood_name,bike_type)


select d.year, neighbourhood_name,count(theft_id)
from bike_dimension b,neighbourhood_dimension n, date_dimension d,bike_theft_fact btf
where b.bike_key=btf.bike_key
and n.neighbourhood_key=btf.hood_key
and d.date_key=btf.date_key
group by cube(d.year,neighbourhood_name)
```

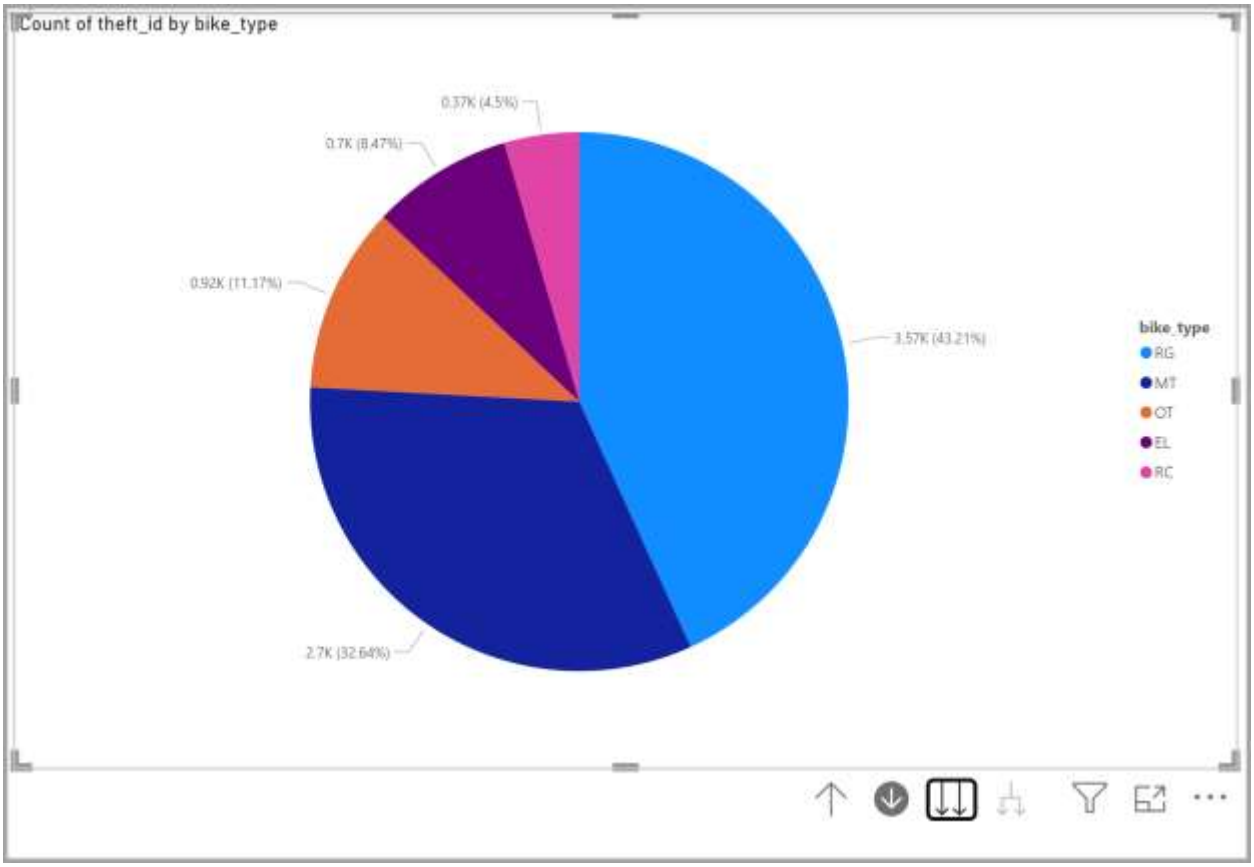
Part II Power BI Reports

Using Power BI desktop to create reports for data visualization to highlight key performance indicators and uncover historical trends. Start Power BI desktop and connect to the database Biketheft. Visualize the reports by opening the file : **bikethetsGrp36_Deliv_D.pbix**.The following reports have been created :

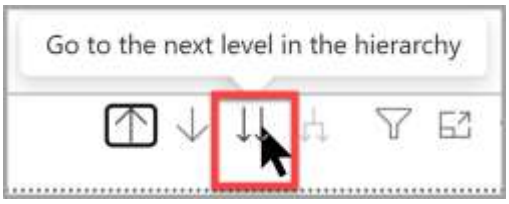
R1.Count bike thefts by year, bike type, neighborhood name

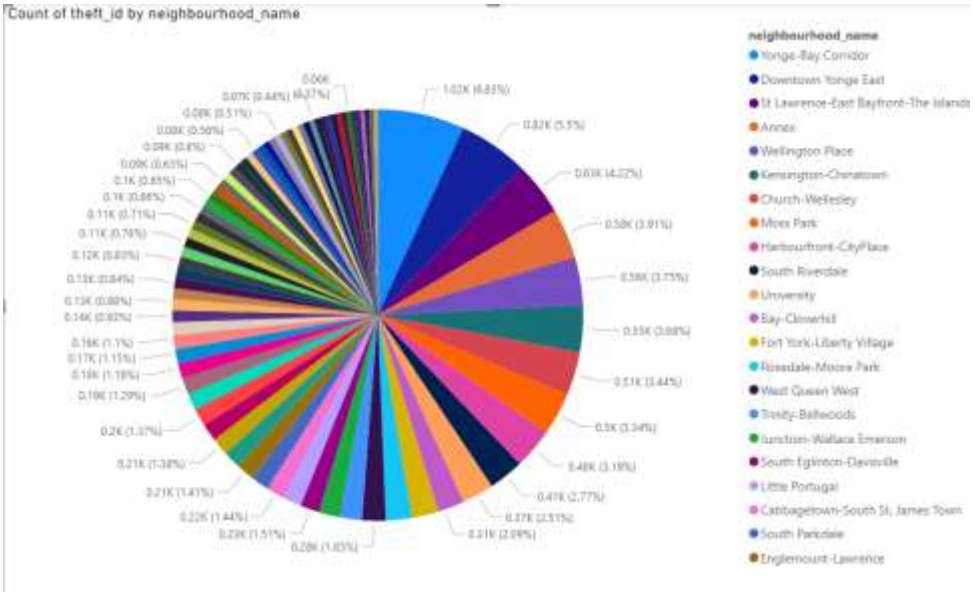




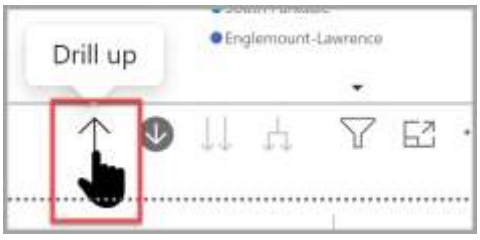


"Go to the next level in the hierarchy" to see by counts by bike type or by neighborhood name:

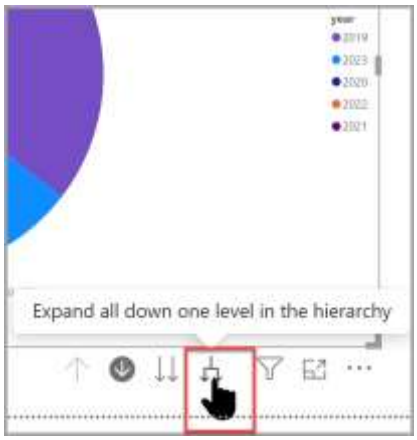


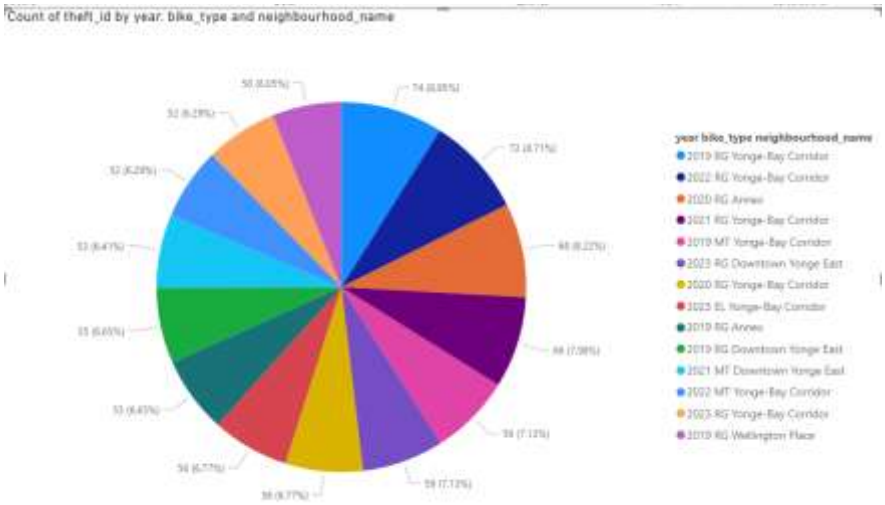


drill up two times:

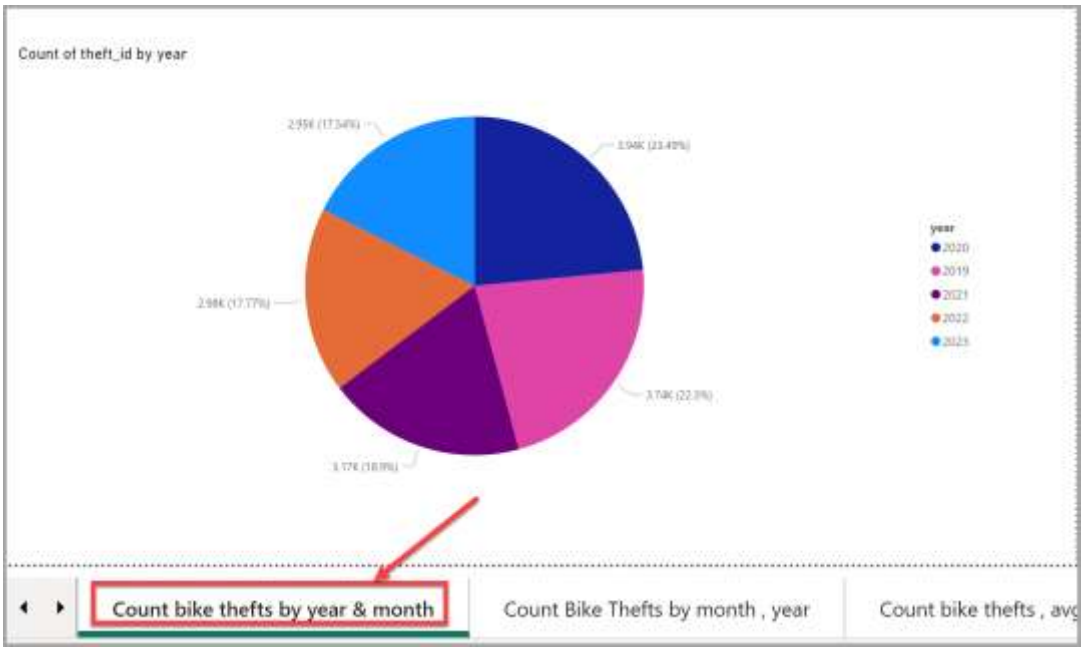


Click on “Expand all down one level in the hierarchy” to see counts by year and bike type and neighborhood name:

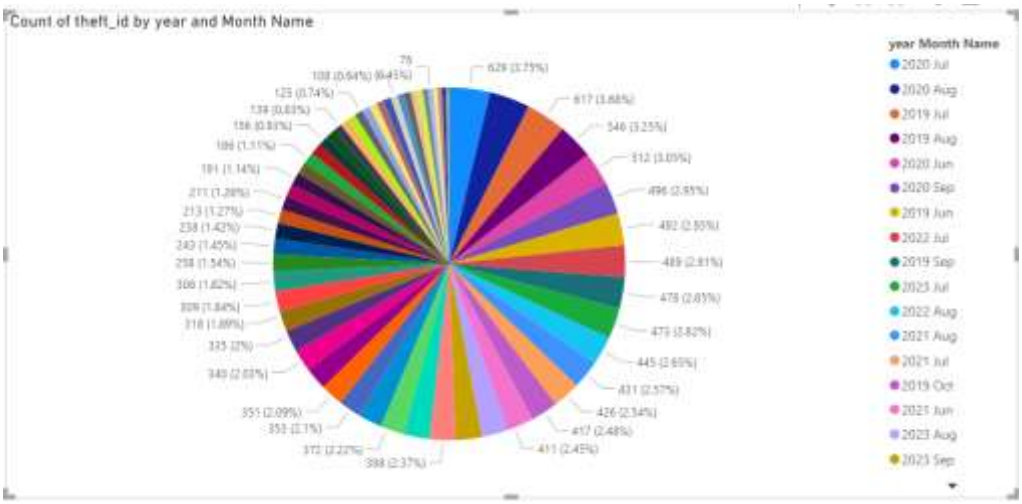




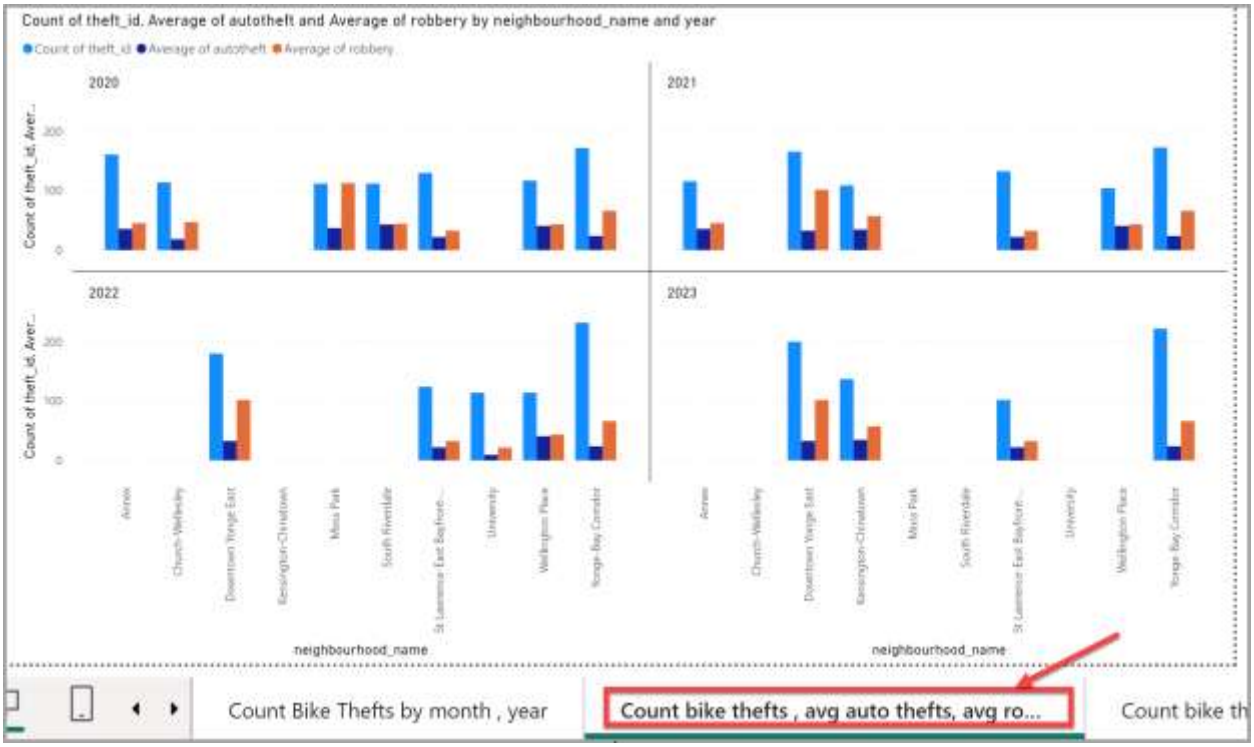
R2 Count thefts by year, month only for the last 5 years:



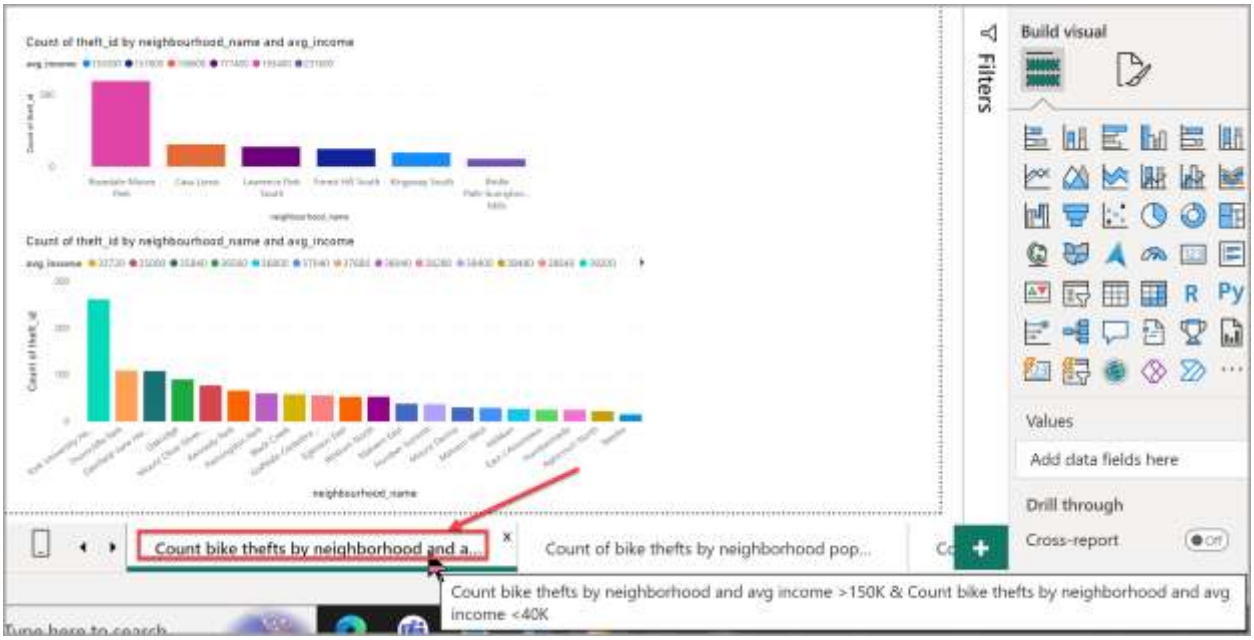
Click on “Expand all down one level in the hierarchy” to see :



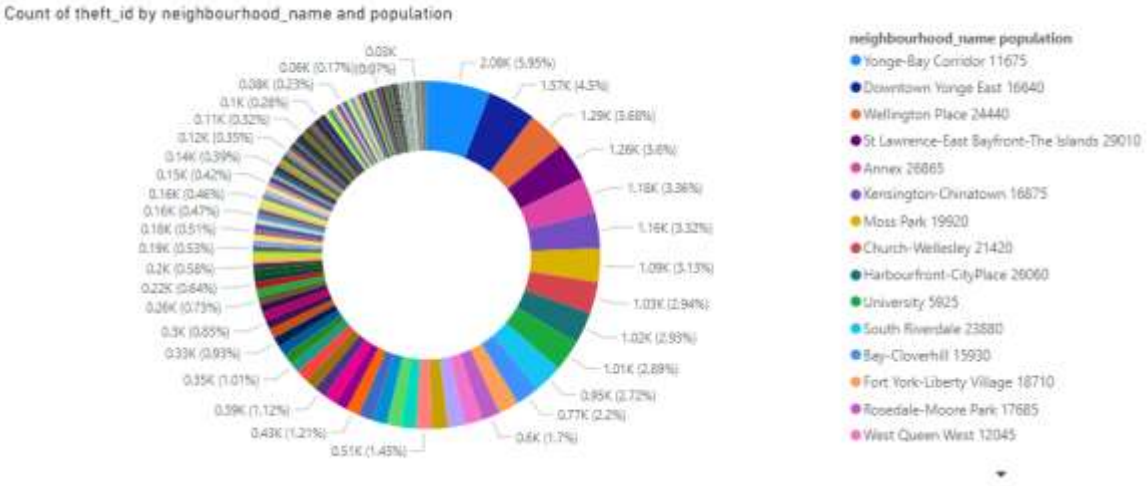
R3. Count bike thefts , avg auto thefts, avg robbery for the last 5 years.



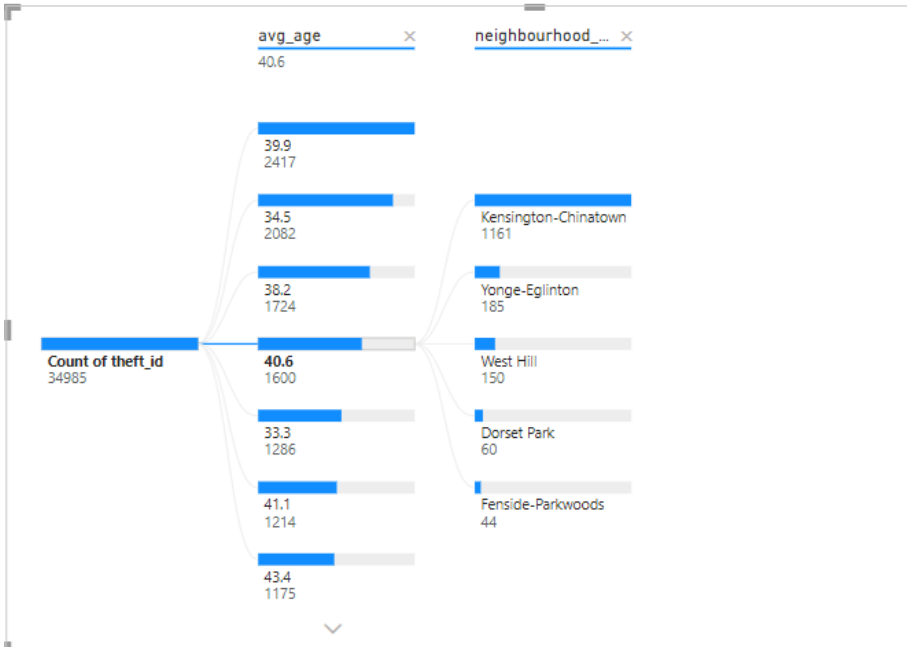
R4. Count bike thefts by neighborhood and avg income >150K & Count bike thefts by neighborhood and avg income <40K.



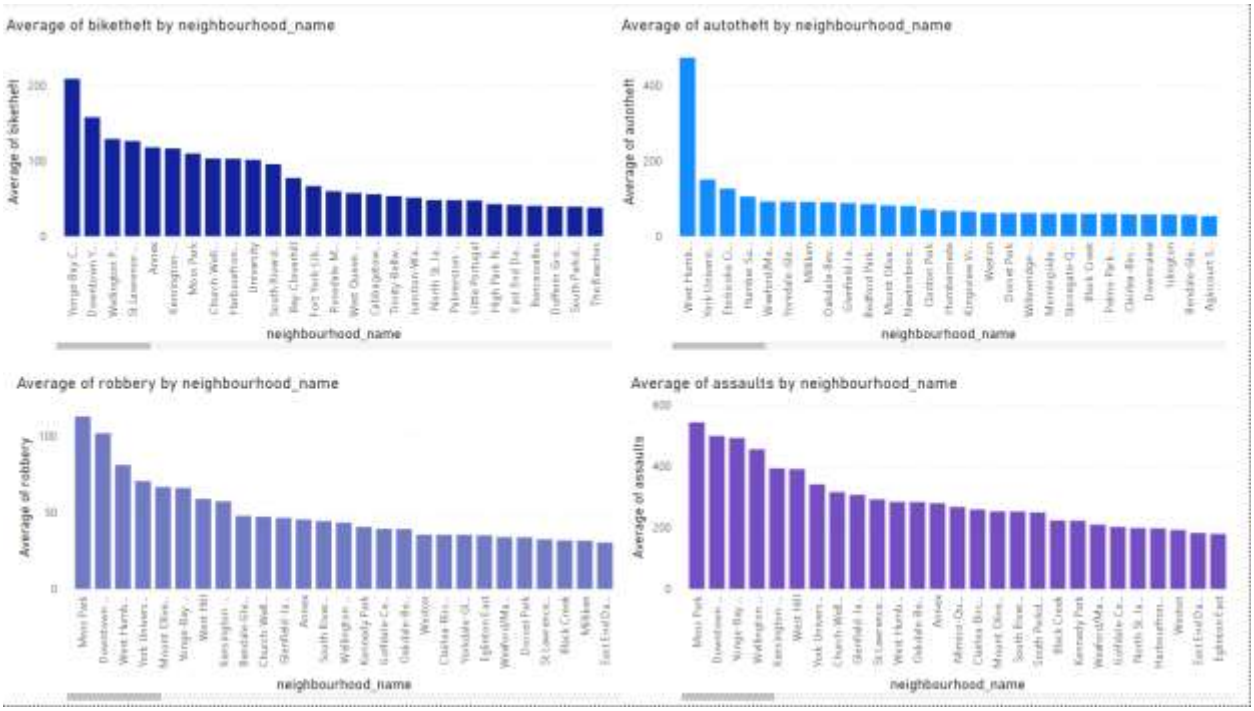
R5 Count of bike thefts by neighborhood population



R6 Count of bike thefts by neighborhood avg age.



R7 Count of bike thefts by neighborhood avg age.



R8 Dashboard (most relevant reports)