

Insights from Denton Crime and Traffic Data Using Static and Streaming Sources

Presented By Team 4:

- Pavan kola
- Abhiram Karthik Meduri
- Abhiram Nallamothe
- Jacob Larkin

CASE STUDY

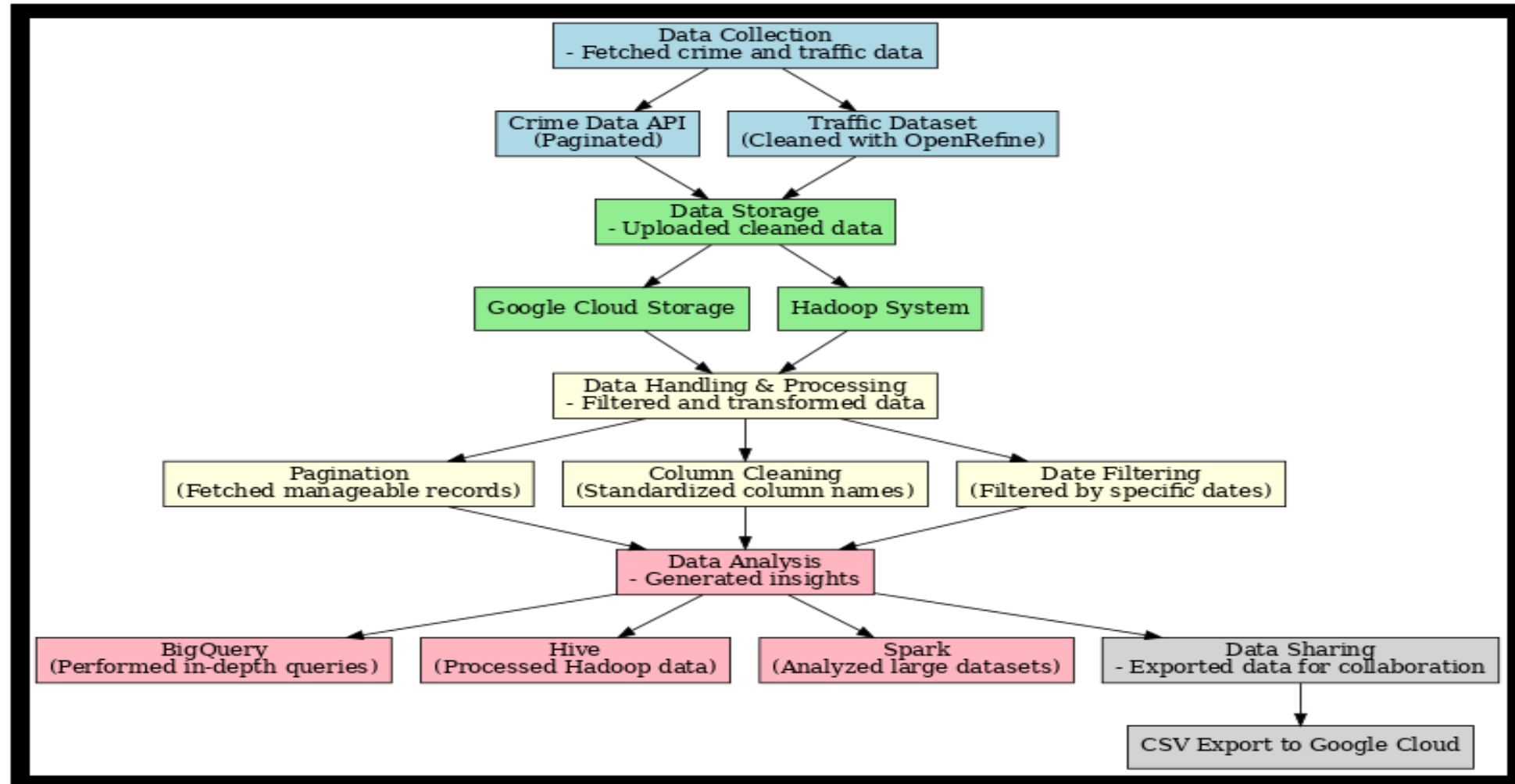
Denton's challenges:

- **Crime**: Increasing property and violent crimes threaten the safety of residents.
- **Traffic**: Rising traffic violations lead to accidents, disruptions, and strain on resources.

Our mission is to turn these datasets into actionable insights to achieve these goals.

- Crime prevention
- Improved traffic enforcement
- Holistic public safety
- Resource optimization

DATA ARCHITECTURE



DATA GENERATION AND COLLECTION

DYNAMIC DATASET

- Generated dynamically through Denton city API.
- Collected from the city of Denton website.
- <https://data.cityofdenton.com/dataset/denton-crime-data>

STATIC DATASET

- Historic data that is captured, it contains traffic violations.
- Collected from the city of Denton website.
- <https://data.cityofdenton.com/dataset/traffic-closed-cases/resource/b12cba00-24e4-4f45-89f8-ffd1c0bf6b95>

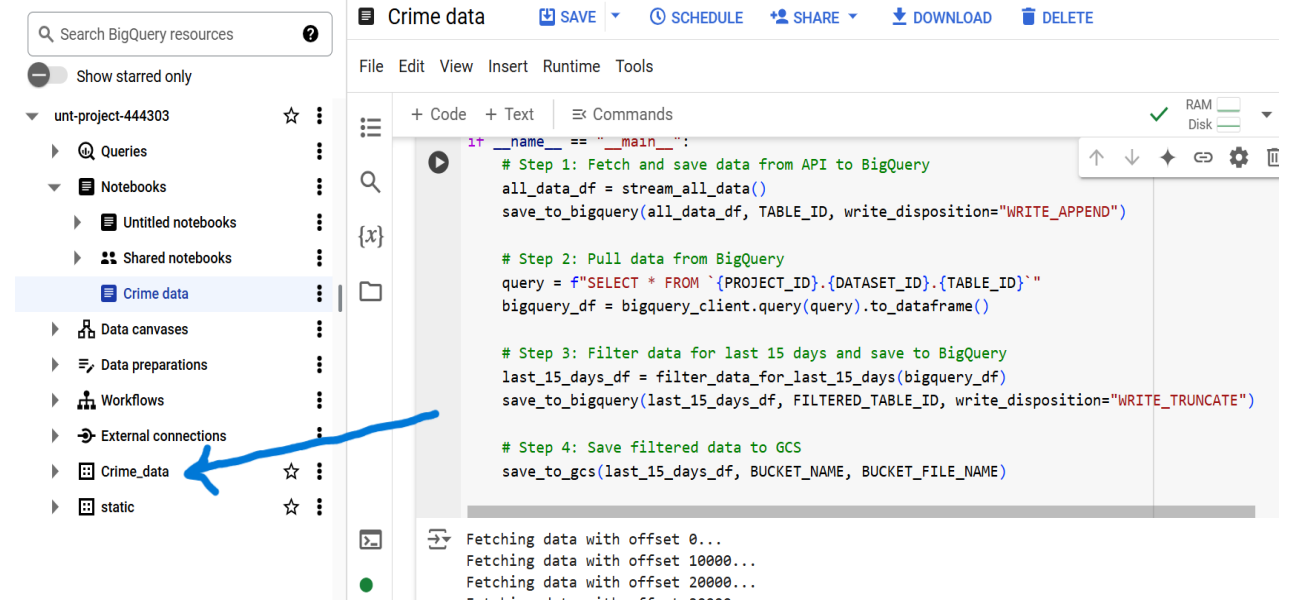
DATA COLLECTION ACTIONS

Crime Dataset (Dynamic)

- Sourced from the City of Denton API to BigQuery.
- Provides real-time access to detailed crime records.

Traffic Dataset (Static)

- Historical traffic violation data provided as a CSV file.
- Manually uploaded to Google Cloud Storage for centralized processing.



The screenshot displays the Google Cloud Platform interface. On the left, a sidebar shows the project 'unt-project-444303' with a tree view containing 'Queries', 'Notebooks', 'Data canvases', 'Data preparations', 'Workflows', 'External connections', 'Crime_data', and 'static'. The 'Crime_data' notebook is selected and highlighted. A blue arrow points from this notebook to the main editor area. The main editor shows the 'Crime data' notebook with a code editor and a console. The code editor contains the following Python code:

```
if __name__ == "__main__":  
    # Step 1: Fetch and save data from API to BigQuery  
    all_data_df = stream_all_data()  
    save_to_bigquery(all_data_df, TABLE_ID, write_disposition="WRITE_APPEND")  
  
    # Step 2: Pull data from BigQuery  
    query = f"SELECT * FROM `{PROJECT_ID}.{DATASET_ID}.{TABLE_ID}`"  
    bigquery_df = bigquery_client.query(query).to_dataframe()  
  
    # Step 3: Filter data for last 15 days and save to BigQuery  
    last_15_days_df = filter_data_for_last_15_days(bigquery_df)  
    save_to_bigquery(last_15_days_df, FILTERED_TABLE_ID, write_disposition="WRITE_TRUNCATE")  
  
    # Step 4: Save filtered data to GCS  
    save_to_gcs(last_15_days_df, BUCKET_NAME, BUCKET_FILE_NAME)
```

The console output shows the following logs:

```
Fetching data with offset 0...  
Fetching data with offset 10000...  
Fetching data with offset 20000...  
Fetching data with offset 30000...
```

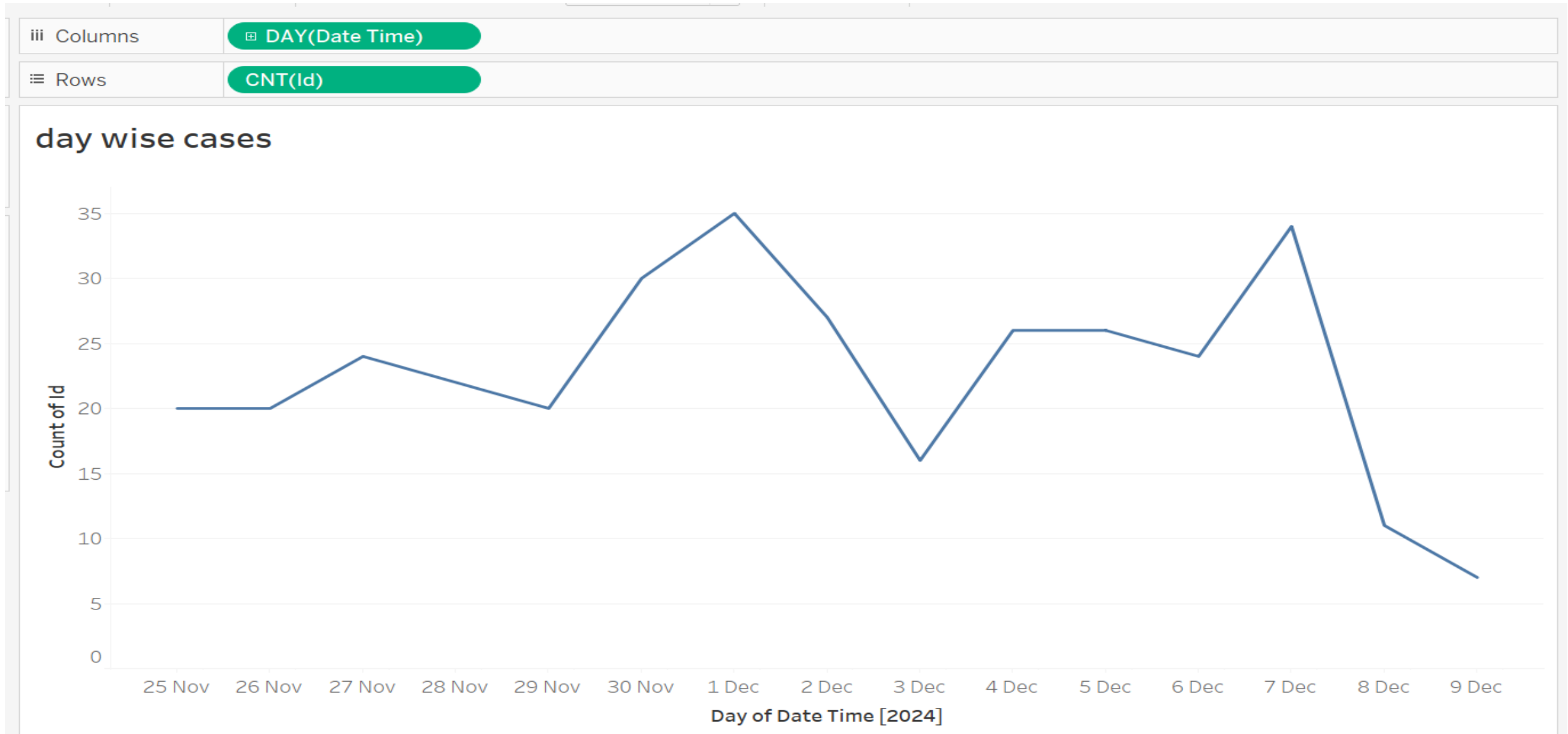
DATA PROCESSING

Step	Description	Tools Used
Cleaning of data	To remove errors and find missing values in the dataset. Fixing them is done	OpenRefine
Transforming of Data	We change formats to make difficult to easily understandable data, which makes to understand the data easily and for analysis too.	SQL,OpenRefine
Executing Queries	We have executed queries to ask different questions and derive meaningful insights	Hadoop, Hive, Spark, BigQuery

MISSING VALUES = 0

TRANSFORMING DATA = Transformed the date column from string to date. Since, some cells of the data column are in different format, we brought every cell into the same format to make the data standardised.

- The formatted column produced the following line chart (JUST REFERENCE):



DATA STORAGE

- **Static dataset:** The data is stored in the google cloud storage bucket, it is stored in the .csv format. Since, applications like hive won't run for other extensions like .xlsx
- **Dynamic dataset:** Fetched the data via the city of Denton API, and then stored in the form of a table in the bigquery for compatibility with querying and analytic tools, also we store the pulled data in the gcp bucket.

The image displays two screenshots from the Google Cloud platform. The top screenshot shows the 'Bucket details' page for a bucket named 'f22abcd' in the 'us-south1 (Dallas)' region. The 'OBJECTS' tab is active, showing a folder browser on the left and a table of objects on the right. The table lists a file named 'traffic-closed-cases-september-2-' with a size of 119.1 KB, type 'text/csv', and a creation date of Nov 20, 2024. The bottom screenshot shows the Google BigQuery interface. The left sidebar lists the project 'unt-project-444303' with various resources, including 'Crime data' which is highlighted. The main area shows the 'Crime data' table with a 'Code' tab selected, displaying a Python script. The script includes comments for four steps: fetching and saving data from an API, pulling data from BigQuery, filtering data for the last 15 days, and saving the filtered data to GCS. The script uses the 'stream_all_data()' function, 'bigquery_client.query()', and 'save_to_gcs()' functions. The bottom of the screenshot shows the execution progress with logs indicating data fetching with offsets.

Google Cloud

ADTA5240F22AKM

buck

Free trial status: \$274.29 credit and 31 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

DISMISS ACTIVATE

Cloud Storage

Bucket details

Overview

Buckets

Monitoring

Settings

f22abcd

Location: us-south1 (Dallas)

Storage class: Standard

Public access: Not public

Protection: Soft Delete

OBJECTS CONFIGURATION PERMISSIONS PROTECTION LIFECYCLE OBSERVABILITY INVENTORY REPORTS OPERATIONS

Folder browser

Buckets > f22abcd > data

CREATE FOLDER UPLOAD TRANSFER DATA OTHER SERVICES

Filter by name prefix only Filter objects and folders

Show Live objects only

Name	Size	Type	Created	Storage class	Last modified	Public access	Version history	Encryption
traffic-closed-cases-september-2-	119.1 KB	text/csv	Nov 20, 2024, 4:32:21 PM	Standard	Nov 20, 2024, 4:32:21 PM	Not public	—	Google-managed

Search BigQuery resources

Show starred only

unt-project-444303

- Queries
- Notebooks
 - Untitled notebooks
 - Shared notebooks
- Crime data
- Data canvases
- Data preparations
- Workflows
- External connections
- Crime_data
- static

File Edit View Insert Runtime Tools

+ Code + Text < Commands

```
if __name__ == "__main__":  
    # Step 1: Fetch and save data from API to BigQuery  
    all_data_df = stream_all_data()  
    save_to_bigquery(all_data_df, TABLE_ID, write_disposition="WRITE_APPEND")  
  
    # Step 2: Pull data from BigQuery  
    query = f"SELECT * FROM `{PROJECT_ID}`.{DATASET_ID}`.{TABLE_ID}`"  
    bigquery_df = bigquery_client.query(query).to_dataframe()  
  
    # Step 3: Filter data for last 15 days and save to BigQuery  
    last_15_days_df = filter_data_for_last_15_days(bigquery_df)  
    save_to_bigquery(last_15_days_df, FILTERED_TABLE_ID, write_disposition="WRITE_TRUNCATE")  
  
    # Step 4: Save filtered data to GCS  
    save_to_gcs(last_15_days_df, BUCKET_NAME, BUCKET_FILE_NAME)
```

Fetching data with offset 0...
Fetching data with offset 10000...
Fetching data with offset 20000...
Fetching data with offset 30000...

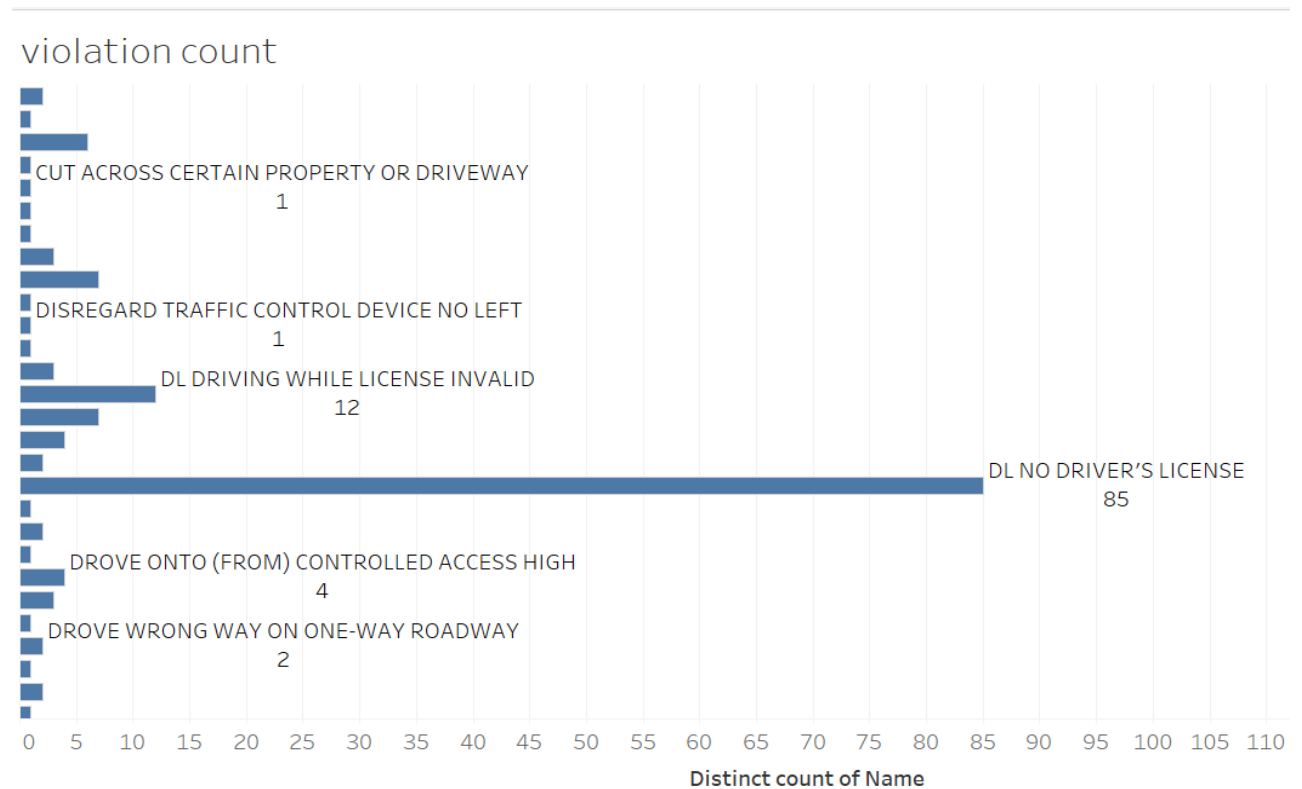
Data Analysis

- Data analysis is done through queries in Hive, spark-sql, BigQuery applications.

Analysis Highlights

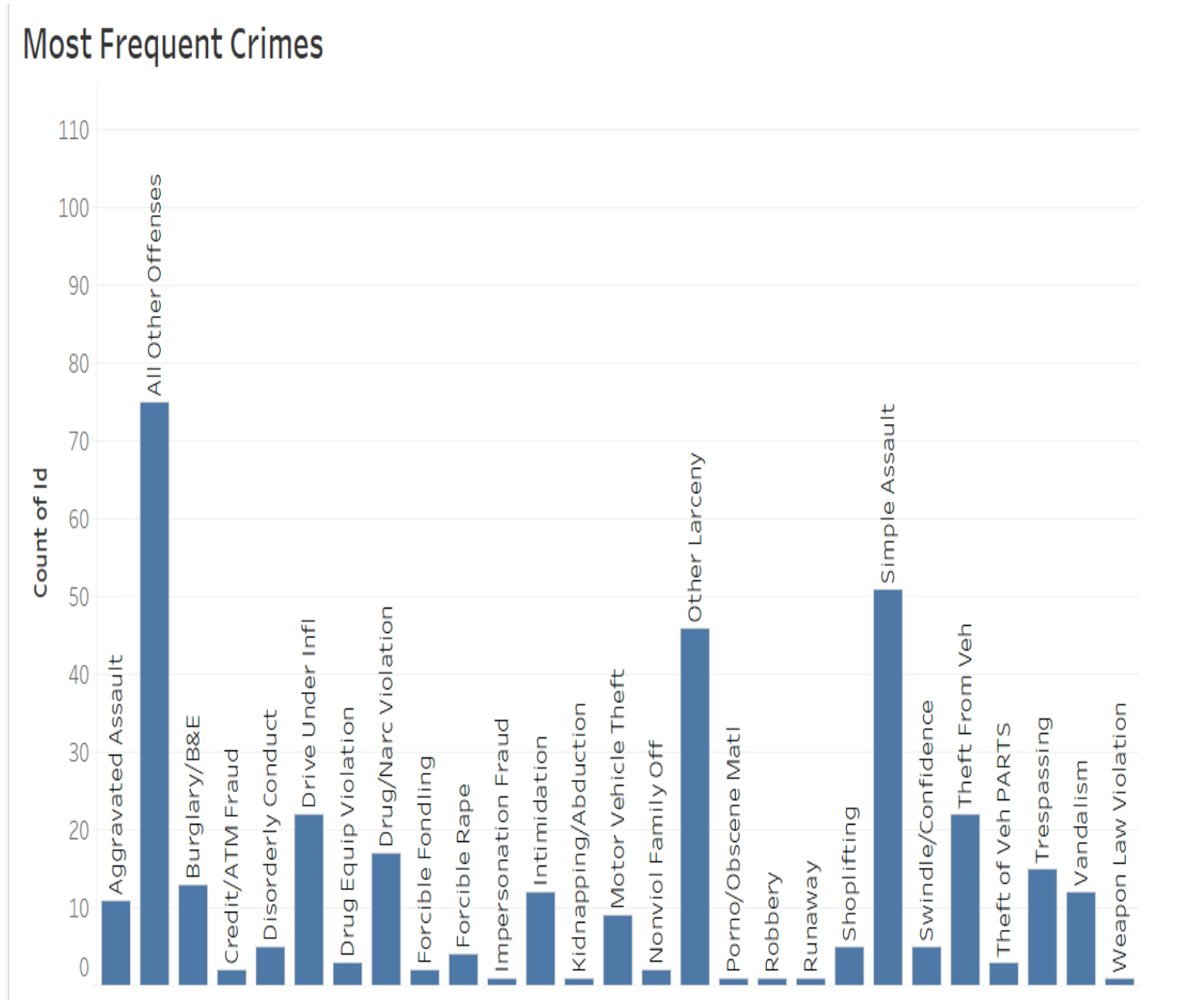
Traffic Data:

- **Frequent Violations:** The most common offenses include driving without a license and invalid license
- **High-Violation Areas:** Teasley Lane and South Loop 288 are hotspots for traffic violations.
- **Violation Trends:** Rush hours (8-10 AM and 4-6 PM) see the highest number of traffic violations.

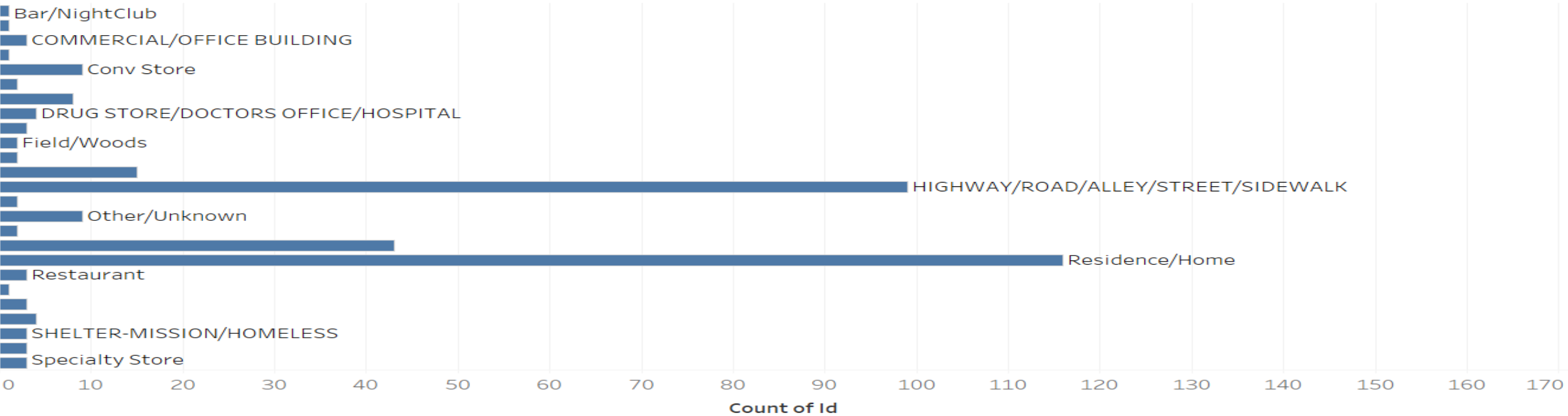


Crime Data:

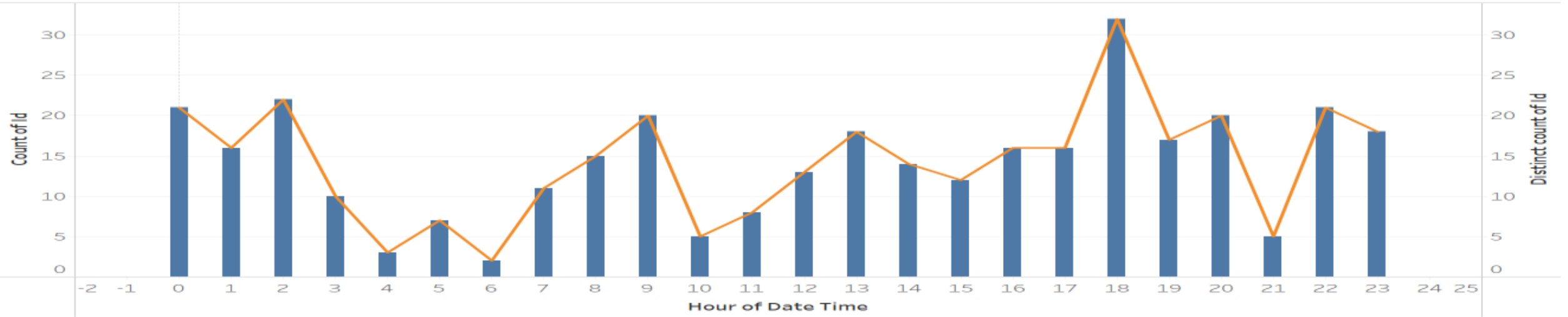
- **Most Common Crimes:** Assault is the most common crime, also there are significant theft crimes like theft from vehicles, motor vehicle theft and theft of vehicle parts.
- **Hotspots:** Crimes are majorly taking place at Residence/Homes, followed by highways/road/alley/sidewalks in the Denton city.
- **Time of the day:** Crimes tend to spike during the evening time at 6:00 pm.



number of crimes by location



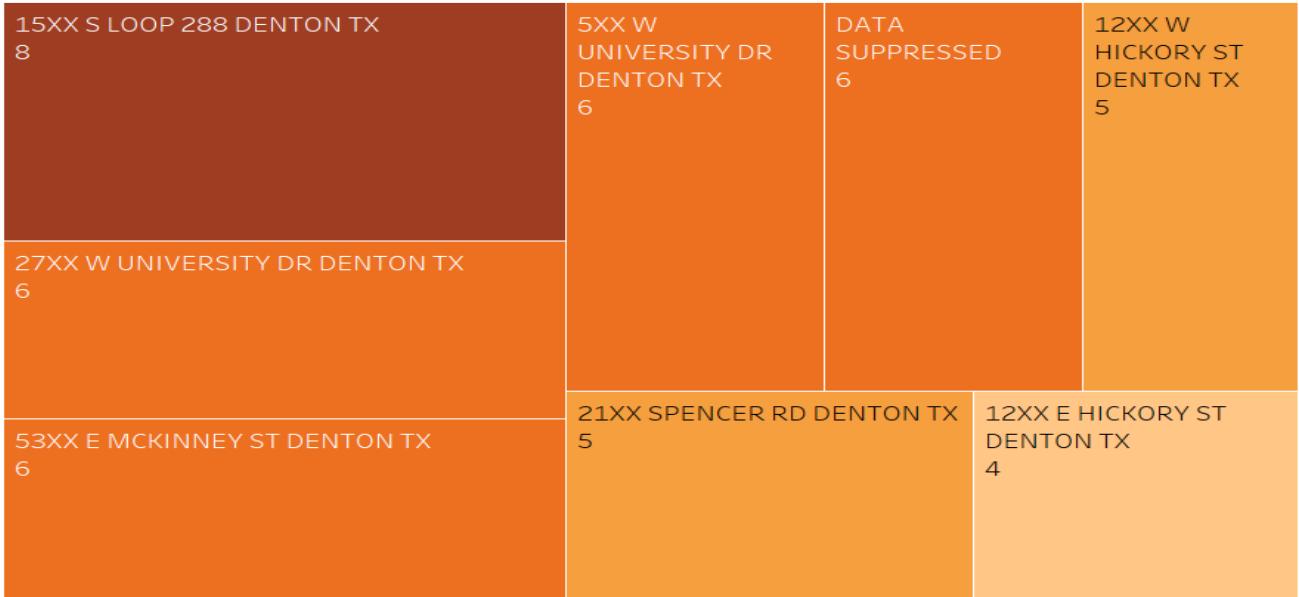
time at which most crimes occur



Data visualizations

- We performed visualizations in Tableau in order to address the most **number of crimes** and the **location** where most crimes are taking place and the **traffic violation hotspots**.
- These help the police to patrol mostly at those places and stop them from being happening.

Denton places with most crime incidents



This visualization helps us to find the exact location in the Denton city, where high crimes are taking place.

The data here is from dynamic dataset, between November 25 2024 to December 7 2024, that means these are related to the very recent crimes that took place in Denton city.

CRIME DATASET THAT INCLUDES POSSIBILITIES OF TRAFFIC VIOLATION

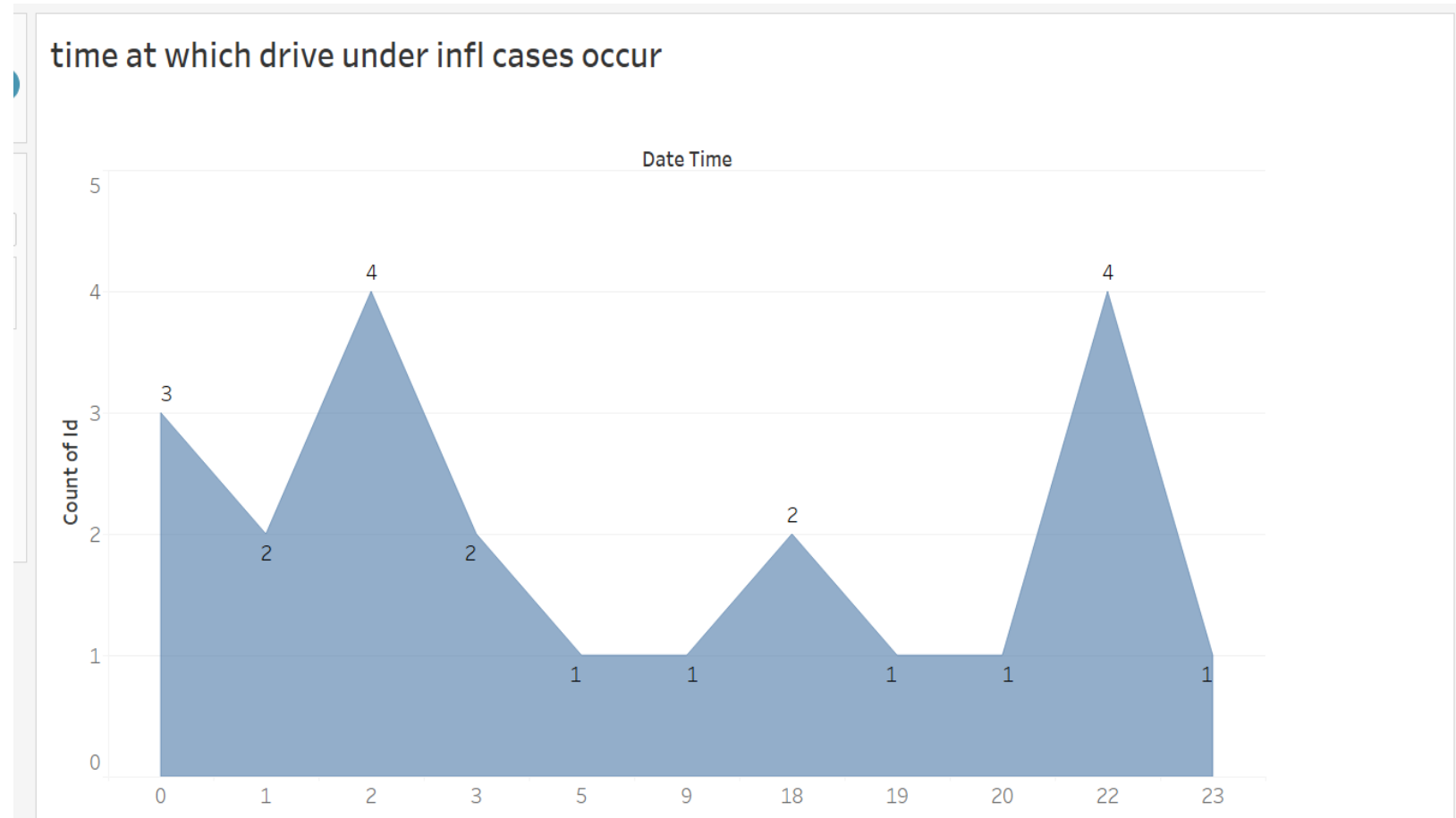
This is the data from then dynamic dataset, where this can lead to the violations of traffic that we are discussing in static dataset.

Driver under infl : driver under influence that means someone who is operating a vehicle while impaired with drugs or alcohol

So if a person is under infl, he is likely to cause traffic violation.

Most of these violations are taking at 10:00pm and 2:00am.

From this we can say that people drink at night are committing to more crimes.



INTERPRETATION / CONCLUSION

Crime Data:

- Most Common Crimes: Assault and theft are the most frequently reported crimes, showing the need for targeted interventions in these areas.
- Hotspots: Crimes are concentrated in residential areas and major roads, highlighting the importance of surveillance in these zones.
- Time Trends: Evening hours, especially around 6:00 PM, see a significant spike in crimes, suggesting the need for increased patrols during these times.

Traffic Data:

- Frequent Violations: Driving without a valid license is one of the most common violations, indicating gaps in driver compliance and enforcement.
- High-Violation Areas: Teasley Lane and South Loop 288 are key traffic violation hotspots, requiring focused enforcement efforts.
- Violation Timing: Most violations occur during rush hours (8-10 AM and 4-6 PM), aligning with higher traffic density.

Overlap Between Crime and Traffic Data:

- Late-night DUI offenses often link traffic violations with crimes, peaking between 10:00 PM and 2:00 AM.
- Overlapping hotspots suggest that combined enforcement strategies could address both issues effectively.

Conclusion/ recommendation for police

- Allocate resources to high-risk areas during peak hours.
- Educate residents on crime prevention and traffic rules.
- Install surveillance systems in key locations to deter violations and crimes.

REFERENCES

1. Practising DatScy (Aug 9,2023), streaming data to a BigQuery table with GCP, Medium, <https://medium.com/@j622amilah/streaming-data-to-a-bigquery-table-with-gcp-240e213829d8>
2. 3 June 2020, What is Tableau and its Importance in Data Visualization?, geeks for geeks, <https://www.geeksforgeeks.org/what-is-tableau-and-its-importance-in-data-visualization/>
3. Cameron Hashemi-pour(2024), what is data life cycle?, TechTarget, <https://www.techtarget.com/whatis/definition/data-life-cycle>
4. Samadritha Ghosh (22nd august 2023), A comprehensive guide to data pre processing, Neptune.ai, <https://neptune.ai/blog/data-preprocessing-guide>

AI slide:

Slide 3 (Data architecture) is AI related slide

THANK YOU