<https://www.linkedin.com/in/francesco-prete-639770153/>



**[Francesco Prete](https://www.linkedin.com/in/francesco-prete-639770153/overlay/about-this-profile/)**

Associate Consultant

Through the Google Data Analytics Professional Certificate I have learned the six steps of data analysis process and I had the opportunity to practice technical tools as Google Sheet, SQL, R and Tableau.

At the end of the course I decided to represent a fictional case study scenario based on an existing public dataset for my Capstone project. The case study analyses the cases of theft at apartment buildings in the city of Chicago for the year 2023. In particular, it will be considered for this study the period between Christmas and New Year’s Eve, in order to discover more precisely whether such cases are more frequent than the average.

As already mentioned, a public dataset will be used in this case study. The structured language SQL will be used for data analysis and Tableau for data visualization.

The choice of using SQL is strictly related to the fact that I want to strengthen my skills with this language not only because I really enjoy playing with data but also because I would like to run queries that could identify high-risk customers in AML and KYC processes, pull transactions data, identify suspicious activities, assessing risk and automate parts of the investigation process.

This Capstone Project will be available on Kaggle and git.hub

I will be following all the data analysis steps: Ask, Prepare, Process, Analyze, Share and Act.

**Scenario:**

You are a junior data analyst working for a business intelligence consultant. You have been at your job for six months, and your boss feels you are ready for more responsibility. He has asked you to lead a project for a brand new client, INTERSEC, a European leader company who provides services to intergovernmental organizations from all over the world.

The company is interested in improving public security in big cities and decided to observe the cases of theft at apartments.

For the team it will be crucial to understand at what time of the year we can see an increase of such cases and elaborate some possible scenarios especially considering the Christmas period and ensure whether during this time cases of theft are below or above the average.

Chicago, the main city of the American state Illinois, was designed for this study.

From these insights the team will design strategies to reduce the frequency of theft at apartments at Christmas time. Ideas, suggestions and potential initiatives will be discussed to the local Police of Chicago. A massive deployment of social media and proposal for investments on public security will be the focus of discussion.

This is intended as a pilot study. In the future the company intends to extend the analysis per city districts and evaluate the situation in other periods of the year. On the basis of the quality of the service provided, the City of Chicago will decide whether to continue the cooperation with INTERSEC.

**ASK**

1. Define a general scenario for the types of crime in Chicago for the year 2023, useful for further studies.
2. How many cases of thefts in apartments has been registered during year 2023 on the average?
3. For the desired period taken into exam – Christmas and New year’s Eve time – which are the trends regarding theft in apartment buildings in Chicago?
4. Proposal of ideas to booster the INTERSEC’s marketing campaign supporting the city of Chicago.

The director of Marketing has assigned me the third and the fourth questions to answer.

Stakeholders are:

-The director of marketing

-The analytics team

My audience is:

-Data Analysts

-Stakeholders

-Citizens of Chicago

I will be providing the key findings with visualizations and recommendations to answer the business case or, more precisely, before the preparation of the data analysis, I will first extract the data I need from the public dataset previously stored on Google Big Query Sandbox. Then I will use SQL for preparing and analyzing the data.

After that, for the representation of the collected data I will use Spreadsheets and BI tool Tableau. All the vizualizations I created are stored in Tableau Public and can be accessed by clicking the link provided right before the vizualizations in this Capstone project.

**PREPARE**

To avoid plagiarism, it is important to mention that for this case study it will be used the following public dataset, *Chicago Crime Data*

https://console.cloud.google.com/bigquery(cameo:product/city-of-chicago-public-data/chicago-crime)?project=bellabeatproj-430920

**A close up of a logo

Description automatically generated**

Type: Data

Category: Public Safety

Dataset origin: Chicago Data Portal <https://data.cityofchicago.org/>

Coloud Service: BigQuery

Made public by: City of Chicago

Update Frequency: Daily

Downloaded data and stored it appropriately on personal Google Big Query Sandbox account.

The entire database consider data from the year 2001 to 2024.

The whole table contains 22 columns.

To make sure that the dataset is reliable it is important to follow first the ROCCC process. Data is reliable, original, comprehensive, current and cited, so it is possible to continue the analysis.

**PROCESS**

To uncover the data and get a general understanding it is necessary to explore it and use some filters:

#1First, I wanted to understand how many cases of crime are committed yearly in Chicago in year 2023

SELECT COUNT (unique\_key) AS total\_cases FROM `bigquery-public-data.chicago\_crime.crime`

WHERE year = 2023;

A screenshot of a computer

Description automatically generated

Result: 262,431

#2How many different types of crime in Chicago for the year 2023?

SELECT COUNT (DISTINCT primary\_type) AS types\_of\_crime FROM `bigquery-public-data.chicago\_crime.crime`

WHERE year = 2023;

A screenshot of a computer

Description automatically generated

Result: 31

#3which types of crimes are the most committed for the year 2023 TOP 5:

SELECT COUNT (unique\_key), primary\_type AS top5 FROM `bigquery-public-data.chicago\_crime.crime`

WHERE year = 2023

GROUP BY (primary\_type)

ORDER BY COUNT(unique\_key) DESC

LIMIT 5;

A screenshot of a computer

Description automatically generated

Result: Theft (57,433), Battery (44,203), Criminal Damage (30,076), Motor Vehicle Theft (29,246), Assault (22,615)

On the basis of these preliminary information for the year 2023 262,419 crime attempts have been committed. Out of the 31 different crime types, theft is the most attempted, followed by battery and then criminal damage, motor vehicle theft and assault.

At this point it is possible to concentrate on the analysis of the crime of theft. To know more about the location on which the cases of theft are perpetrated:

#4in which places have been registered cases of theft, top5:

SELECT COUNT(unique\_key), location\_description FROM `bigquery-public-data.chicago\_crime.crime`

WHERE primary\_type = 'THEFT'

AND year = 2023

GROUP BY(location\_description)

ORDER BY COUNT (unique\_key) DESC

LIMIT 5;

A screenshot of a computer

Description automatically generated

Result: Street (15,742), Apartment (7,355), Small Retail Store (4,393), Residence, Department Store

Most of the cases are registered on the street, in apartments and stores. On the basis of these general results we know for now that for the year 2023, 7,355 cases of theft have been committed at apartment buildings.

**ANALYZE**

Once collected the part of data necessary for the analysis we can start the examination of the selected information.

#1 It will be useful to discover of the average of such cases weekly for the considered year.

SELECT COUNT (unique\_key) / 52.0 AS avg\_weekly\_23 FROM `bigquery-public-data.chicago\_crime.crime`

WHERE primary\_type = 'THEFT'

AND location\_description = 'APARTMENT'

AND date BETWEEN '2023-01-01 00:00:00 UTC' AND '2023-12-31 23:59:59 UTC'

A screenshot of a computer

Description automatically generated

Result: 141.44

On average, we know that more than 140 cases of theft in the apartments are registered in Chicago weekly.

The weekly average will be important for the next part of the analysis, as it will be compared to the weeks under Christmas period.

We need to discover whether during the weeks containing the Christmas day and the New Year’s Eve the cases of theft at apartments are below or above the average and possibly understanding the reasons.

#2Cases of theft in apartments under the Christmas period for the weeks:

1. **2023-12-18 – 2023-12-24**
2. **2023-12-25 – 2023-12-31**
3. **2024-01-01 – 2024-01-01**

SELECT COUNT (unique\_key) AS week1

FROM `bigquery-public-data.chicago\_crime.crime`

WHERE primary\_type = 'THEFT'

AND location\_description = 'APARTMENT'

AND date BETWEEN '2023-12-18 00:00:00 UTC' AND '2023-12-24 23:59:59 UTC'

A screenshot of a computer

Description automatically generated

Result: 184

SELECT COUNT (unique\_key) AS week2

FROM `bigquery-public-data.chicago\_crime.crime`

WHERE primary\_type = 'THEFT'

AND location\_description = 'APARTMENT'

AND date BETWEEN '2023-12-25 00:00:00 UTC' AND '2023-12-31 23:59:59 UTC'

A screenshot of a computer

Description automatically generated

Result: 143

SELECT COUNT (unique\_key) AS week3

FROM `bigquery-public-data.chicago\_crime.crime`

WHERE primary\_type = 'THEFT'

AND location\_description = 'APARTMENT'

AND date BETWEEN '2024-01-01 00:00:00 UTC' AND '2024-01-08 23:59:59 UTC'

A screenshot of a computer

Description automatically generated

Result: 147

Given that the weekly average of theft crimes in apartment in Chicago was 141.44 for the whole year 2023, we can observe that the week 1, the one right before the Christmas day shows a far above the average rate of theft at apartments, whereas week 2 and week 3 that contain the Christmas day and the New Year’s Eve are slightly over the average.

If we consider two more week for the period December 2023 it is possible to notice that both of them are as well over the average.

The week from 2023-12-02 to 2023-12-09 shows as result 177 cases, whereas for the week from 2023-12-10 to 2023-12-17 we count 162 cases.

**A screenshot of a computer

Description automatically generated**

A screenshot of a computer

Description automatically generated

An explanatory chart that compares the cases of thefts in apartments for the weeks in December 2023 and the weekly average for the entire year. Available also on Tableau Public: https://public.tableau.com/app/profile/francesco.prete/viz/Book1\_17250543623520/Sheet1

#3 avg month for year 2023

SELECT COUNT (unique\_key) / 12.0 AS avg\_month\_year23

FROM `bigquery-public-data.chicago\_crime.crime`

WHERE primary\_type = 'THEFT'

AND location\_description = 'APARTMENT'

AND date BETWEEN '2023-01-01 00:00:00 UTC' AND '2023-12-31 23:59:59 UTC'

A screenshot of a computer

Description automatically generated

Result: 612

#4 this query shows the total cases for dec2023:

SELECT COUNT (unique\_key) AS dec\_23

FROM `bigquery-public-data.chicago\_crime.crime`

WHERE primary\_type = 'THEFT'

AND location\_description = 'APARTMENT'

AND date BETWEEN '2023-12-01 00:00:00 UTC' AND '2023-12-31 23:59:59 UTC'

A screenshot of a computer

Description automatically generated

Result: 702

We can conclude that for the period December 2023 the registered cases of theft at apartment buildings are on the whole more than the average, especially in the first half of the month and with a peak on week 2023-12-18 and 2023-12-24 (week1).

Moreover December 2023 is the month on which were registered more than the average cases of theft in apartment buildings.

**SHARE**

Considering the fact that I am not able to use Google Big Query sandbox directly to create visualization nor to create new tables, I have collected the analyzed data into spreadsheets to be uploaded onto Tableau to create the proper visualization.

You will see bar charts, a tree map and line charts regarding:

a) types of crime in Chicago for the year 2023

b) location where cases of thefts are registered

c) theft in apartments in Chicago

d) average month for the year 2023 vs. average December 2023

To use Tableau, it was necessary to upload spreadsheets.

Then by clicking SHEET 1 on the bottom left of the screen you can work drafting your visualization.

You can start working with the information you have on the left part of the screen, Table, on panel Data. On columns I put SUM(f\_0), whereas on Rows, primary\_type.

Then I selected Bar, in the menu Marks.

I got the chart below:

A screenshot of a computer

Description automatically generated

I wanted to choose a tree map, more compelling to represent immediately the five main types of crime in Chicago during the year 2023 and I renamed it *Top5 Crimes in Chicago 2023.*

*https://public.tableau.com/app/profile/francesco.prete/viz/Top5CrimesinChicago/Sheet1*

**a.**

**A blue squares with black text

Description automatically generated**

To do so I clicked on Show me, on the right part of the screen and I chose the tree map chart.

As the data analysis regards the crime of theft, the most committed in the city, I had to show in which places is mostly registered, so I opted for a bar chart. Each location on which theft cases are registered, is marked with a different color and a legenda is shown on the right of the screen.

To do so:

Columns: location\_description

Rows: SUM(f0\_)

Then:

Color: location\_description

Label: primary\_type, SUM(f\_0) <show mark labels>

https://public.tableau.com/app/profile/francesco.prete/viz/Top5TheftLocationsinChicago2023/Sheet1?publish=yes

**b)**

**A graph with different colored squares

Description automatically generated**

For the next table in order to show the situation for a given period with peaks and dips I chose a line chart.

https://public.tableau.com/app/profile/francesco.prete/viz/Book1\_17250543623520/Sheet1?publish=yes

**c)**

**A graph with blue lines

Description automatically generated**

Moreover In this chart I have inserted the average line that shows the average for this month.

To do so, on the right side of the screen: Analytics > Summarize > Average Line > drag it and drop it to the chart.

The next bar chart show the difference between the average of thefts in apartment during December 23 vs. the monthly average for the whole year 2023

https://public.tableau.com/app/profile/francesco.prete/viz/average\_dec23vs\_monthly\_average\_23/Sheet1?publish=yes

**d)**

**A graph of value and value

Description automatically generated**

**ACT**

After showing a general situation in the city for the year 2023, I could reach to these conclusions:

Although the date included in this study don’t show the evidence of this, December 2023 was as well the month in which there has been the highest number of registered cases of theft at apartments, whereas the month on which the lowest number of cases has been registered for year 2023 was February, that is at the same time the month that represents a rate below the monthly average.

Curiously, the climate in December doesn’t seems to have any influence on the frequency of such cases (daily temperatures are between 0°C and 6°C Celsius degrees)

For the month December 2023 the week that registered the most cases was *2023-12-18 – 2023-12-24* (*week 1* for this study). Probably this is the week during which residents are more likely to be away: most people leave their apartments in order to spend the Christmas time with relatives and families or on holiday or just out shopping and that would explain such higher frequency.

For these reasons, I would suggest that during the first half of the month and especially during the days before Christmas it would be better to intensify local police patrol – even undercover - in the city to lower the rate, but also implement surveillance cameras services, invest in lighting and secure entry systems.

Raising awareness: educate residents about theft preventions during Christmas period via posts on social medias, by organizing meetings or distribution of leaflets.

The same study can be further developed in the future considering also other variables such as the city districts and different times of the year.