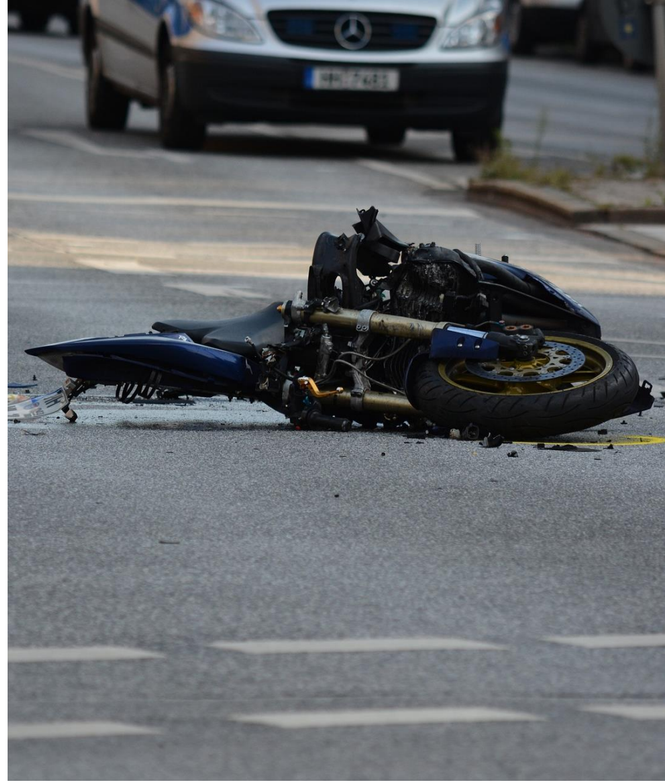
Accident Insights:

Exploring the Potential and Challenges of Storing Accident Data in SQL Relational Database



Presented by: Edem Dziko

Introduction

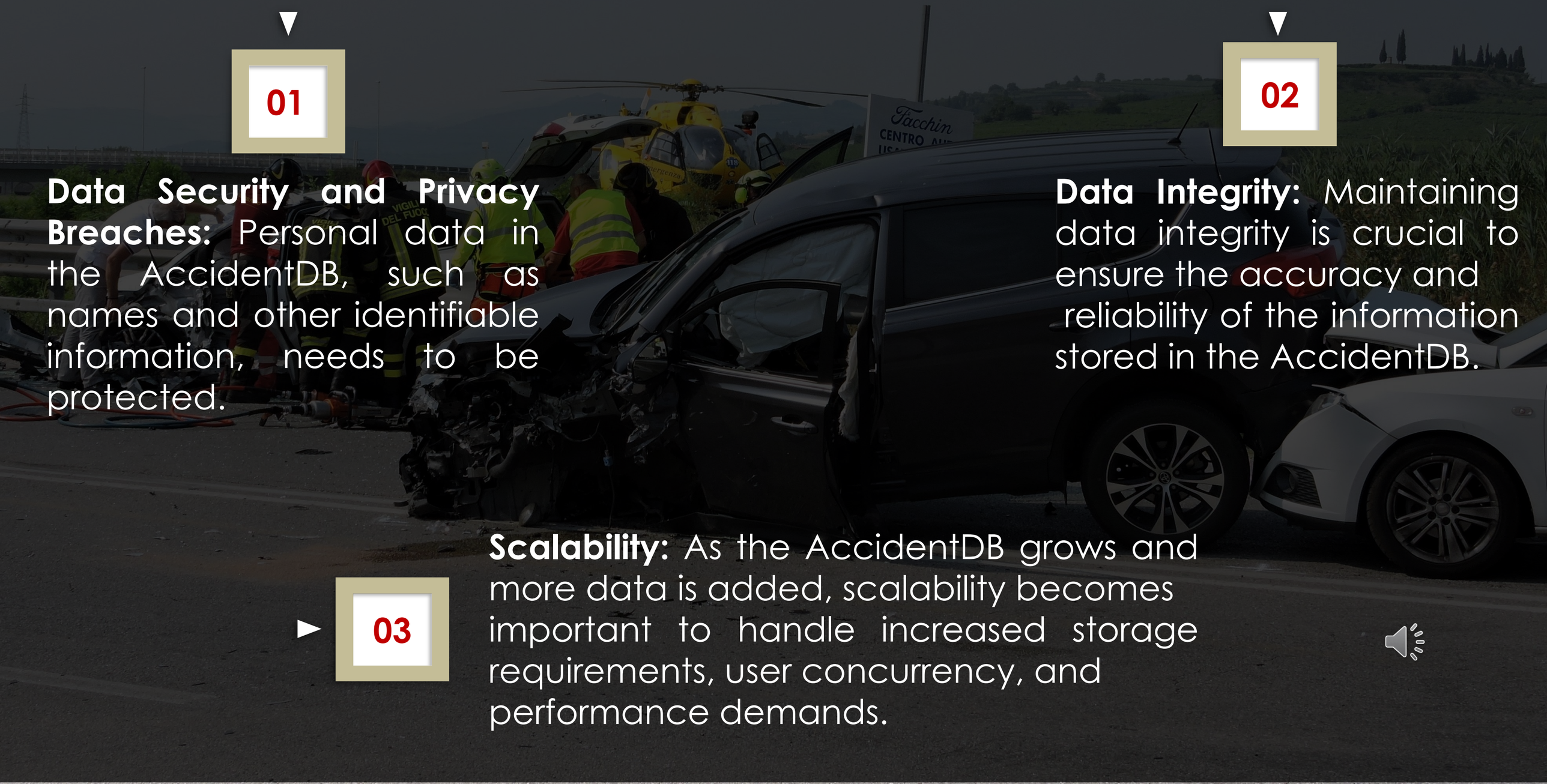
- Road traffic accidents continue to be a major cause of injuries and fatalities, thus becoming an urgent concern for people all over the world (Gururaj., 2008). Governments and organisations work to reduce these risks through a variety of strategies, such as collecting accident data, better infrastructure, stronger traffic laws, and public awareness programmes (WHO., 2015)

- The accident database contains information about accidents in the UK

- The database contains 4 tables namely Accident, Casualty, Vehicle and LSOA

- The aim of this project is to showcase the power of data analysis and the role of a data scientist in leveraging an SQL relational database to gain valuable insights from the accident database

Technical Issues of Storing Data



01

Data Security and Privacy Breaches: Personal data in the AccidentDB, such as names and other identifiable information, needs to be protected.

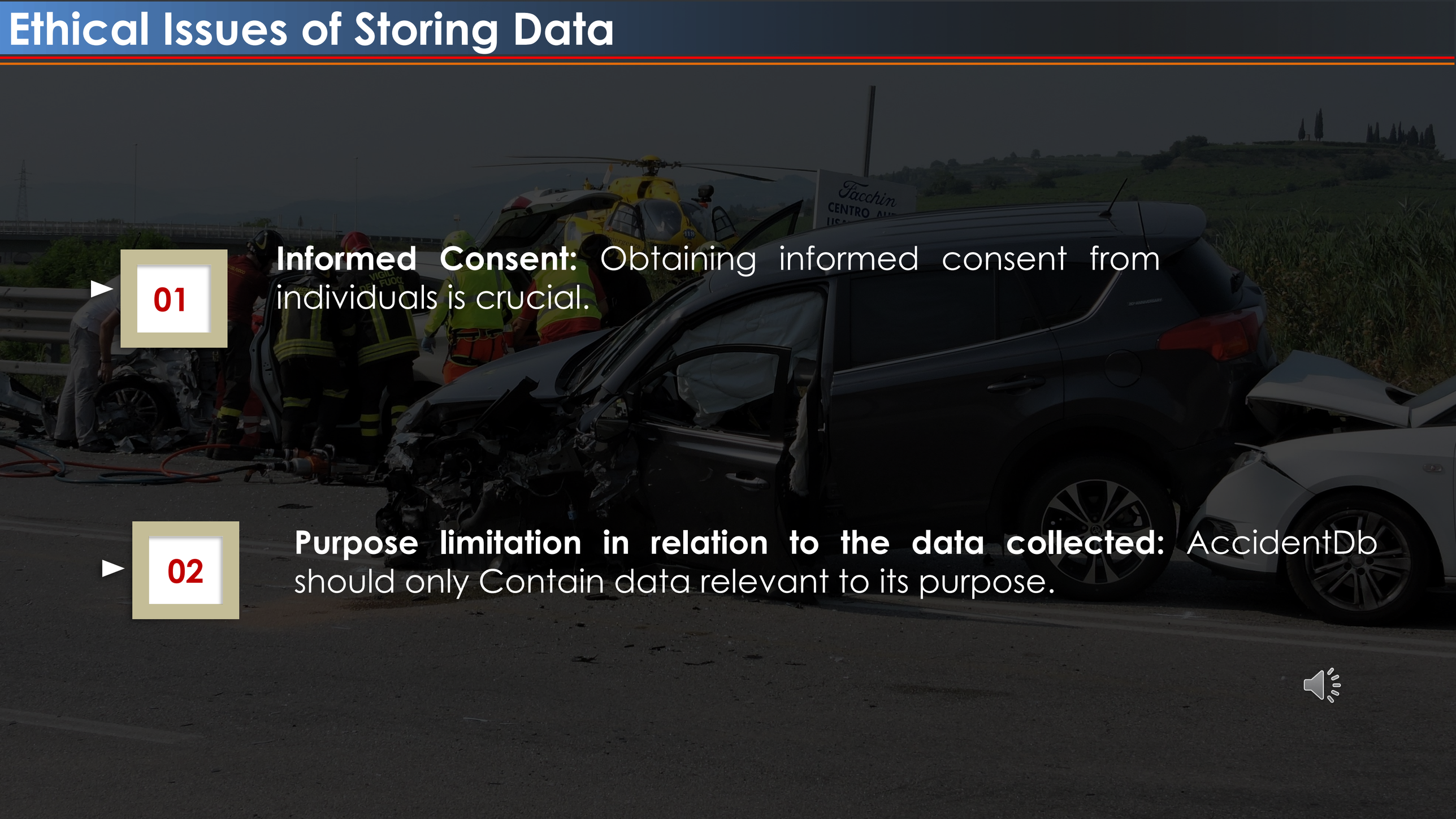
02

Data Integrity: Maintaining data integrity is crucial to ensure the accuracy and reliability of the information stored in the AccidentDB.

03

Scalability: As the AccidentDB grows and more data is added, scalability becomes important to handle increased storage requirements, user concurrency, and performance demands.

Ethical Issues of Storing Data



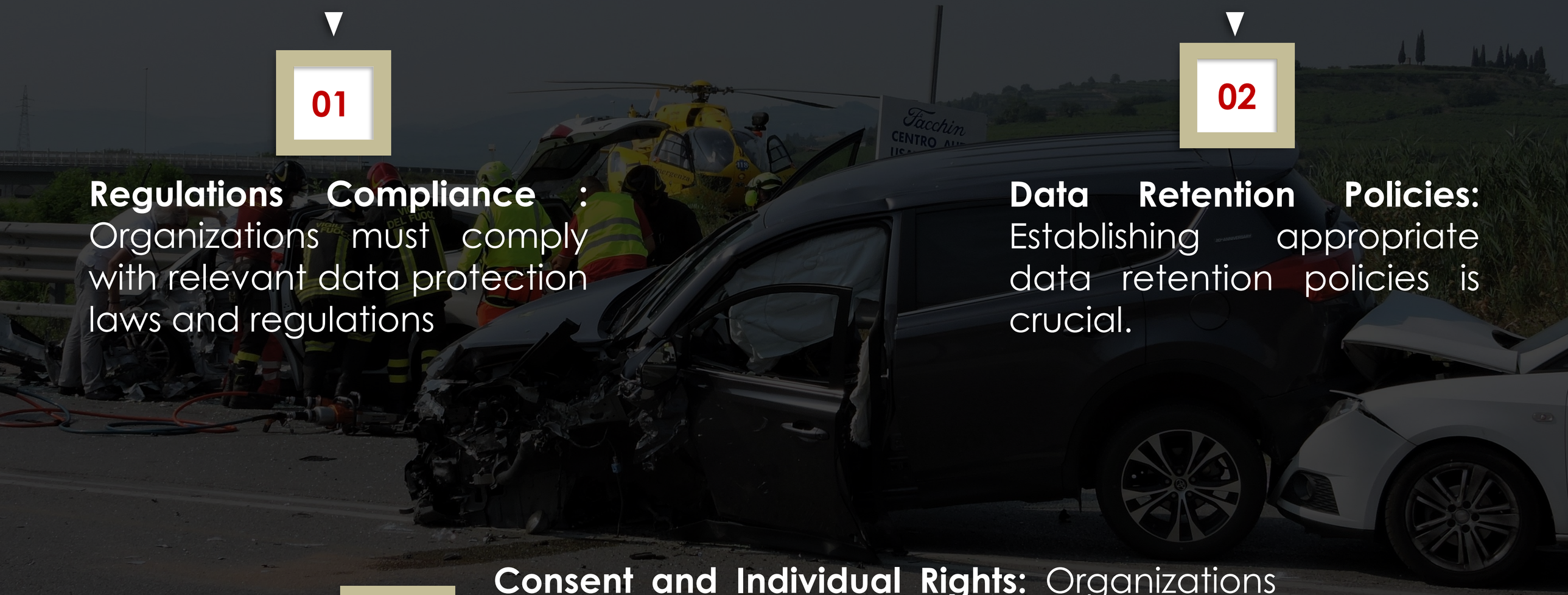
01

Informed Consent: Obtaining informed consent from individuals is crucial.

02

Purpose limitation in relation to the data collected: AccidentDb should only Contain data relevant to its purpose.

Legal Issues of Storing Data



01

Regulations Compliance : Organizations must comply with relevant data protection laws and regulations

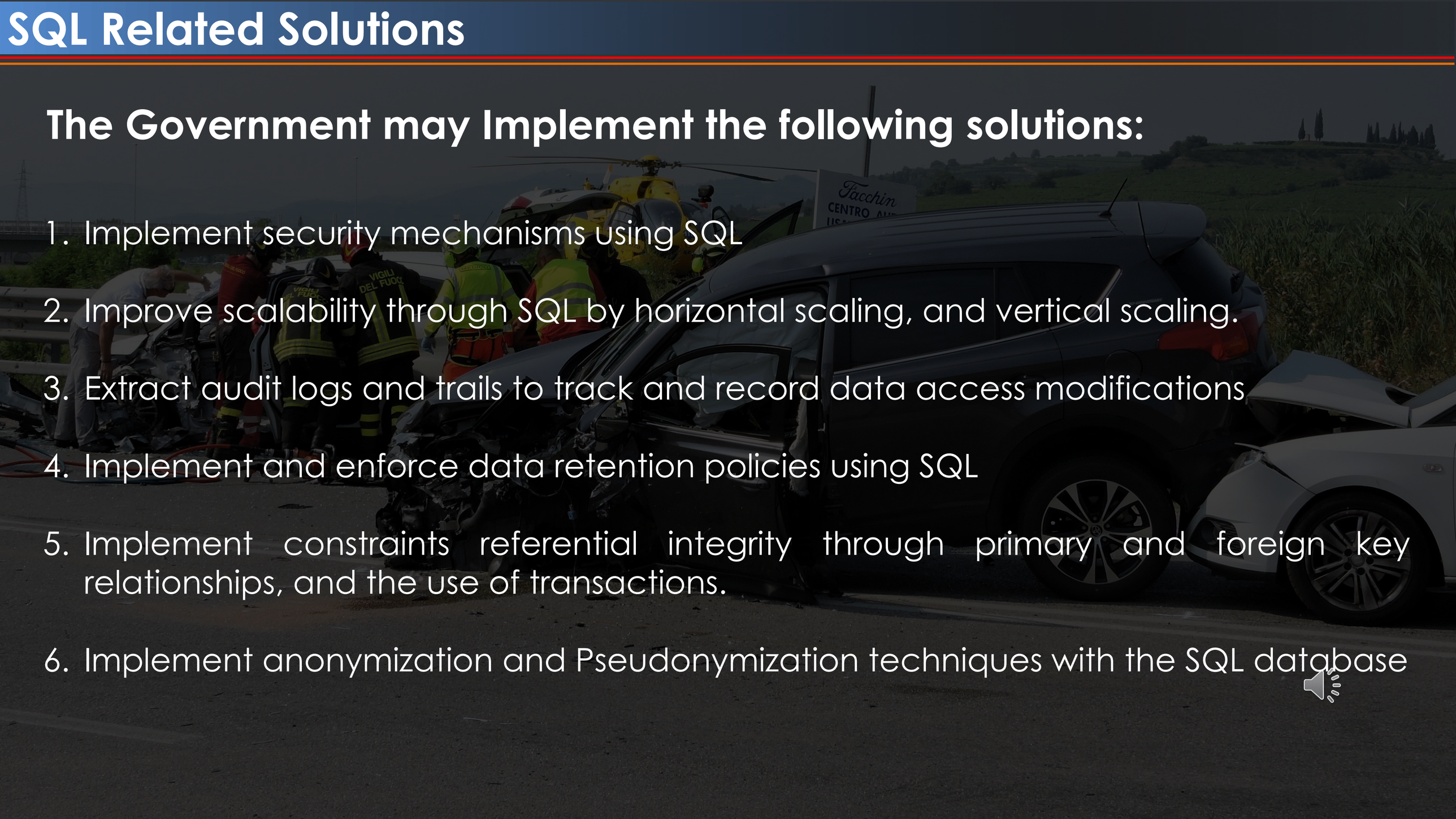
02

Data Retention Policies: Establishing appropriate data retention policies is crucial.

03

Consent and Individual Rights: Organizations must obtain appropriate consent from individuals for the collection, processing, and storage of their personal data.

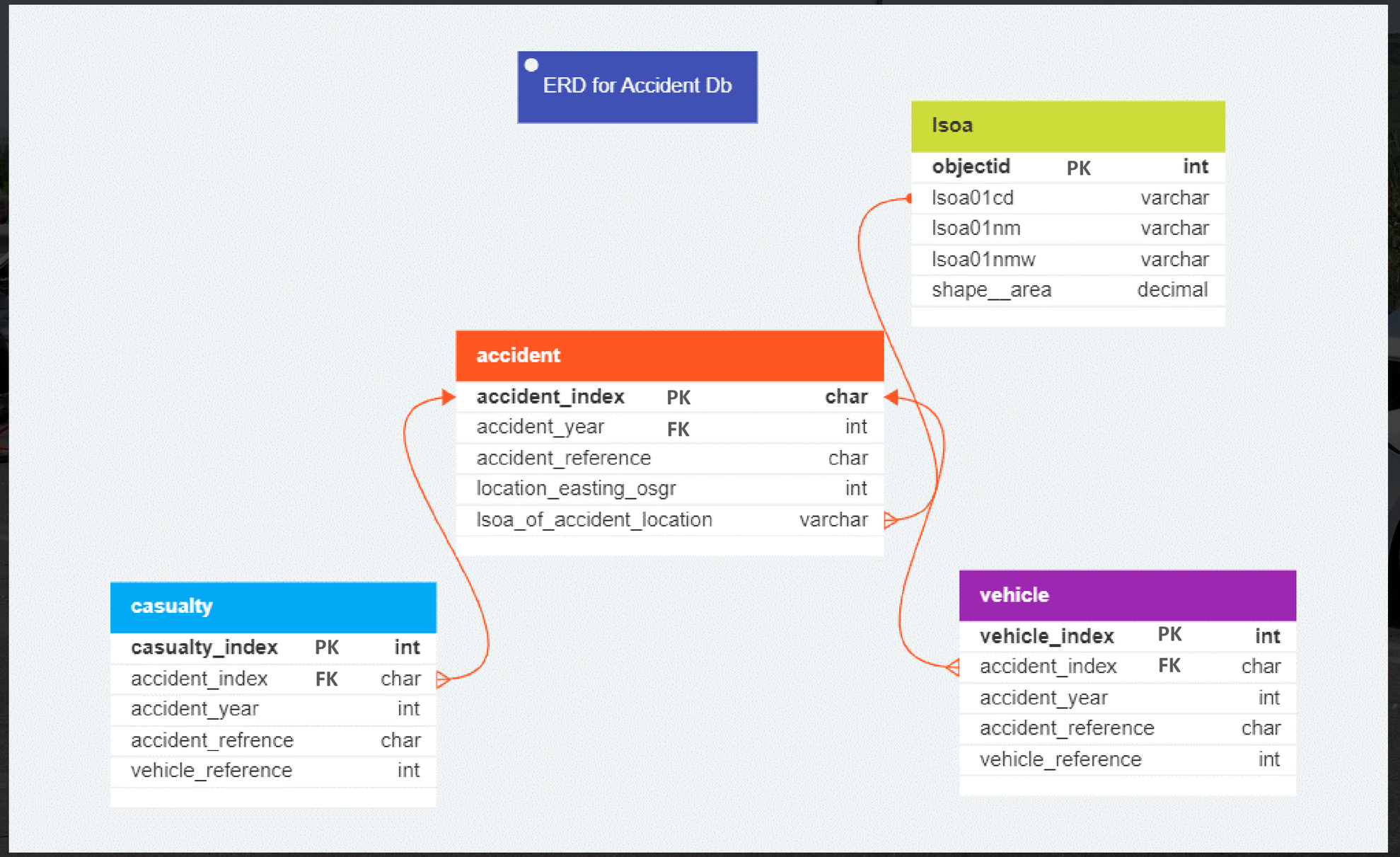
SQL Related Solutions



The Government may Implement the following solutions:

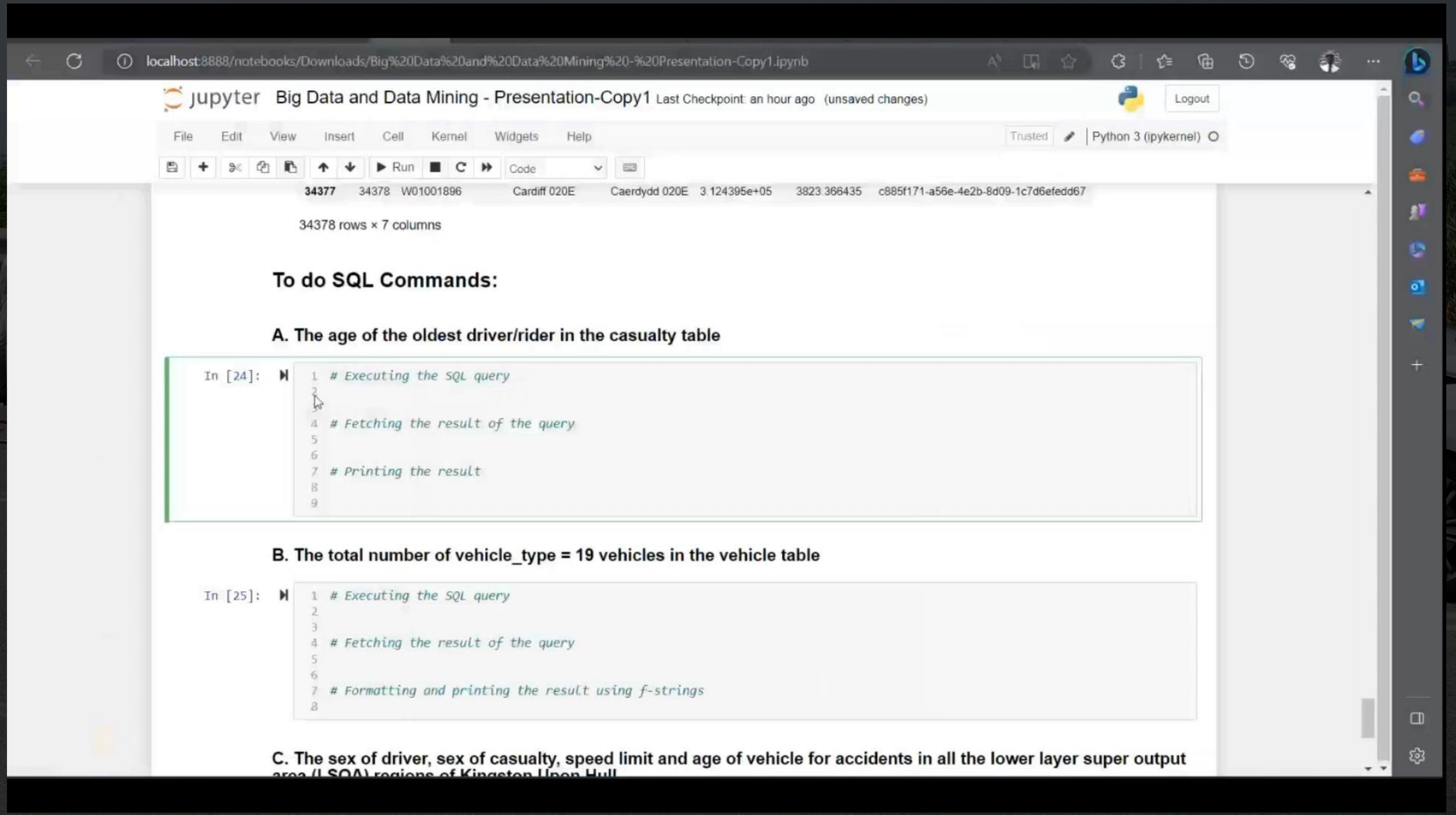
1. Implement security mechanisms using SQL
2. Improve scalability through SQL by horizontal scaling, and vertical scaling.
3. Extract audit logs and trails to track and record data access modifications
4. Implement and enforce data retention policies using SQL
5. Implement constraints referential integrity through primary and foreign key relationships, and the use of transactions.
6. Implement anonymization and Pseudonymization techniques with the SQL database

Entity Relationship Diagram

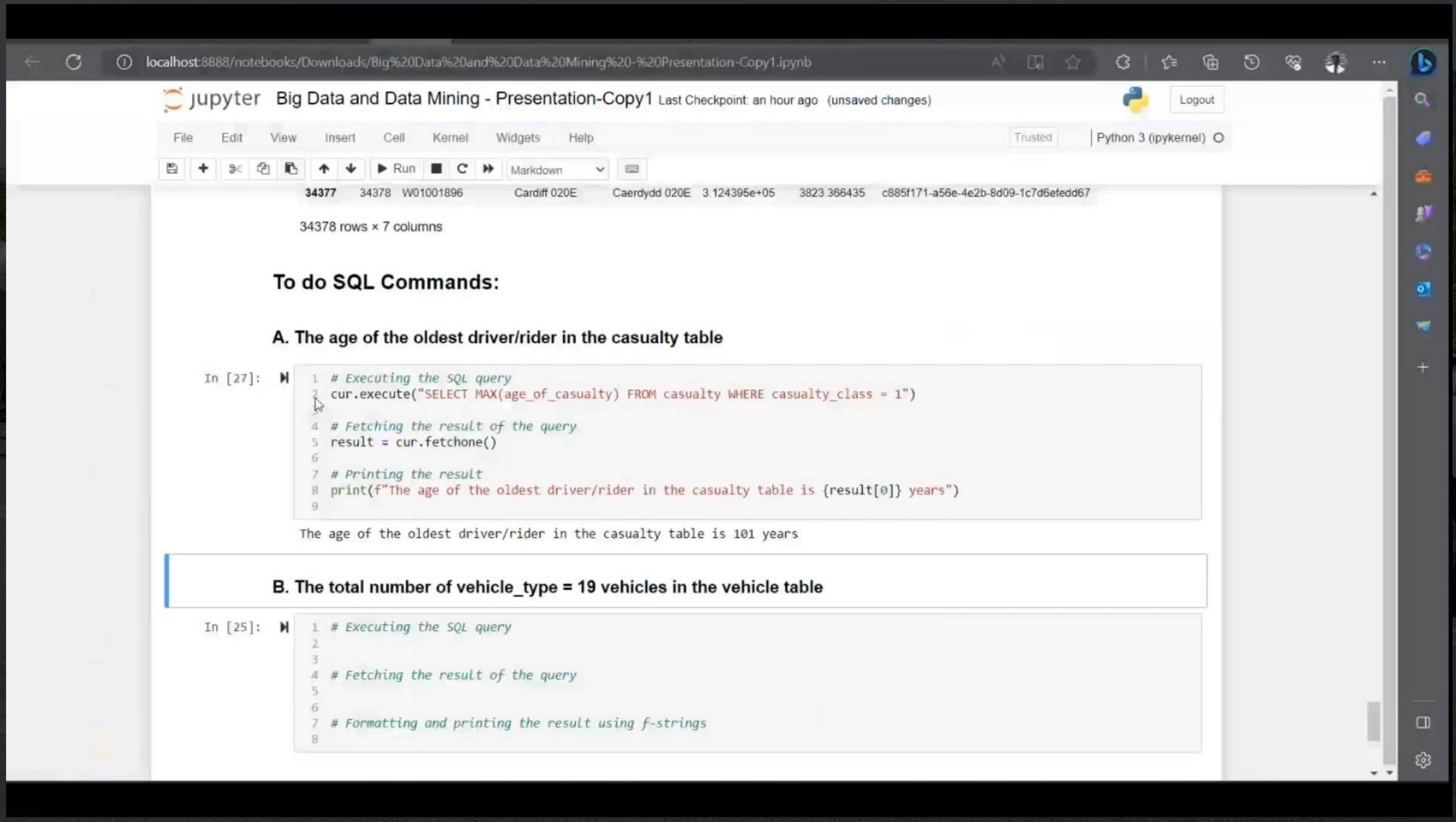


SQL Tutorials

Age of the oldest driver/rider in the casualty table



The Total number of vehicle type= 19 vehicles



The sex of driver, casualty, speed limit & age of vehicles in LSOA of Hull

localhost:8888/notebooks/Downloads/Big%20Data%20and%20Data%20Mining%20-%20Presentation-Copy1.ipynb

jupyter Big Data and Data Mining - Presentation-Copy1 Last Checkpoint: an hour ago (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run Markdown

In [27]:

```
# Executing the SQL query
cur.execute("SELECT MAX(age_of_casualty) FROM casualty WHERE casualty_class = 1")

# Fetching the result of the query
result = cur.fetchone()

# Printing the result
print(f"The age of the oldest driver/rider in the casualty table is {result[0]} years")

```

```
The age of the oldest driver/rider in the casualty table is 101 years
```

B. The total number of vehicle_type = 19 vehicles in the vehicle table

In [29]:

```
# Executing the SQL query
cur.execute("SELECT SUM(vehicle_type =19) AS No_Van_Goods_Vehicle FROM vehicle")

# Fetching the result of the query
result = cur.fetchone()

# Formatting and printing the result using f-strings
print(f"The number of Van/Goods Vehicle are {result[0]}")
```

```
The number of Van/Goods Vehicle are 47458
```

C. The sex of driver, sex of casualty, speed limit and age of vehicle for accidents in all the lower layer super output area (LSOA) regions of Kingston Upon Hull

In [26]:

```
# Executing the SQL query

# Fetching the result of the query and storing it in a pandas DataFrame


# Printing the DataFrame
```

References

Gururaj, G., 2008. Road traffic deaths, injuries and disabilities in India: current scenario. National Medical Journal of India, 21(1), p.14.

World Health Organization, 2015. Global status report on road safety 2015. World Health Organization.

Chandrakar, I. and Hulipalled, V.R., 2022. Improved technique for preserving privacy while mining real time big data. International Journal of Communication Networks and Information Security, 14(1), pp.86-92.

Wagner, J., Rasin, A., Glavic, B., Heart, K., Furst, J., Bressan, L. and Grier, J., 2017. Carving database storage to detect and trace security breaches. Digital Investigation, 22, pp.S127-S136.

Samaraweera, G.D. and Chang, J.M., 2019. Security and privacy implications on database systems in Big Data era: A survey. IEEE Transactions on Knowledge and Data Engineering, 33(1), pp.239-258.

Sicari, S., Rizzardi, A. and Coen-Porisini, A., 2022. Security&privacy issues and challenges in NoSQL databases. Computer Networks, 206, p.108828.