

The background is a blue gradient with decorative white circuit-like lines in the corners. These lines consist of straight segments and small circles, resembling a stylized electronic circuit or data flow diagram.

# CYBER SECURITY BREACH

# ABOUT THE PROJECT

Data and Security Breaches have become a bigger problem for companies in this digital world. Cyber Security plays an important role in tackling with these problems.

The objective of this project is to create a dashboard which can help in tracking breaches happened in the past and events related to those breaches.

# ABOUT THE DATA

The dataset contains various attributes which helps in getting details related to the breaches happened in the past and events related to those breaches. The details about those attributes are :

- Number : Unique Id number for each breach.
- Name of Covered Entity : Name of Company/Organisation which was main target of breach.
- Business Associate Involved : Name of Company/Organisation which was impacted along with the main Company/Organisation.

- Total Individuals : Number of employees working in the Company/Organisation.
- Individuals Affected : Number of employees impacted from the breach.
- Type of Breach : Category of Breach.
- Location of Breach Information : Source Device/Product where breach was detected.
- Breach Start : Date when breach was detected.
- Breach End : Date when issue related to breach was closed. If no date is given then issue is not resolved.
- Branch : Name of branch which was impacted from the breach.
- Department : Name of department which was target of the breach.
- Country Branch : Name of country where Company/Organisation's data was breached.
- Employee( who find out breach ) : Cyber Security Employee who detected the breach.
- Employee URL : Profile Picture URL of the employee who detected the breach.
- Estimated Stolen Data(GB) : Amount of data which was compromised in the breach.

# EDA AND DATA TRANSFORMATION

The steps involved in EDA and Data Transformation are :

- **Checking and Treating Null values** : Dataset which will be used for analysis and dashboard creation must be free from null values. In our dataset, there were zero null values and data was ready for analysis.
- **Column DataType** : In a dataset, each column should have correct datatype, so that analysis can be done in correct way. In our dataset, Breach Start and Breach End columns were changed to Date from text and Estimated Stolen Data(GB) was changed to numeric from text datatype.

- **Creating Various Tables for Data Modelling :** Data modelling is process of analysing the data objects and their relationship to the other objects. The Data Model's main focus is on what data is needed and how we have to organize data rather than what operations we have to perform. So for data modelling, we have to create various tables (known as Dimension Tables) and then establish there relationship with Main Table (known as Fact Table) to create Star Schema. Tables created are Branches, Departments, Breach Types, Countries, Employees who found the breach.

# ARCHITECTURE DESCRIPTION

- Raw data collection : Data is a open source file and collected form Kaggle website.
- Data Pre-processing and Transformation : Before building any model, it is important to perform data processing and transformation to make data suitable for data analysis and visualization. The processes followed in this project are :
  - Checking null values : Dataset was checked for null values and data was free from null values.



- Checking and Correcting datatype of columns : The datatype of each column should be according to the values inside it for better analysis and visualization. The datatype of Columns Breach Start and Breach End was changed from Text to Date and Estimated Stolen Data(GB) from Text to Numeric datatype.

➤ Data Modelling : Data modelling is process of analysing the data objects and their relationship to the other objects. The Data Model's main focus is on what data is needed and how we have to organize data rather than what operations we have to perform. So for data modelling, we have to create various tables (known as Dimension Tables) and then establish there relationship with Main Table (known as Fact Table) to create Star Schema. Tables created are Branches, Departments, Breach Types, Countries, Employees who found the breach.



- Deployment : Dashboards are created using Power BI Desktop and then deployed in workspace at Power BI Web. From Power BI web client can access using their relevant account.

## Dashboards

## Cyber Breaches Dashboard

Home

Cyber Experts

Branch

Department

Country

Total Breaches  
1055

Closed Breaches  
145

Open Breaches  
910

Avg Days Per Closed Breach  
243.63

Estimated Stolen Data(GB)  
138K

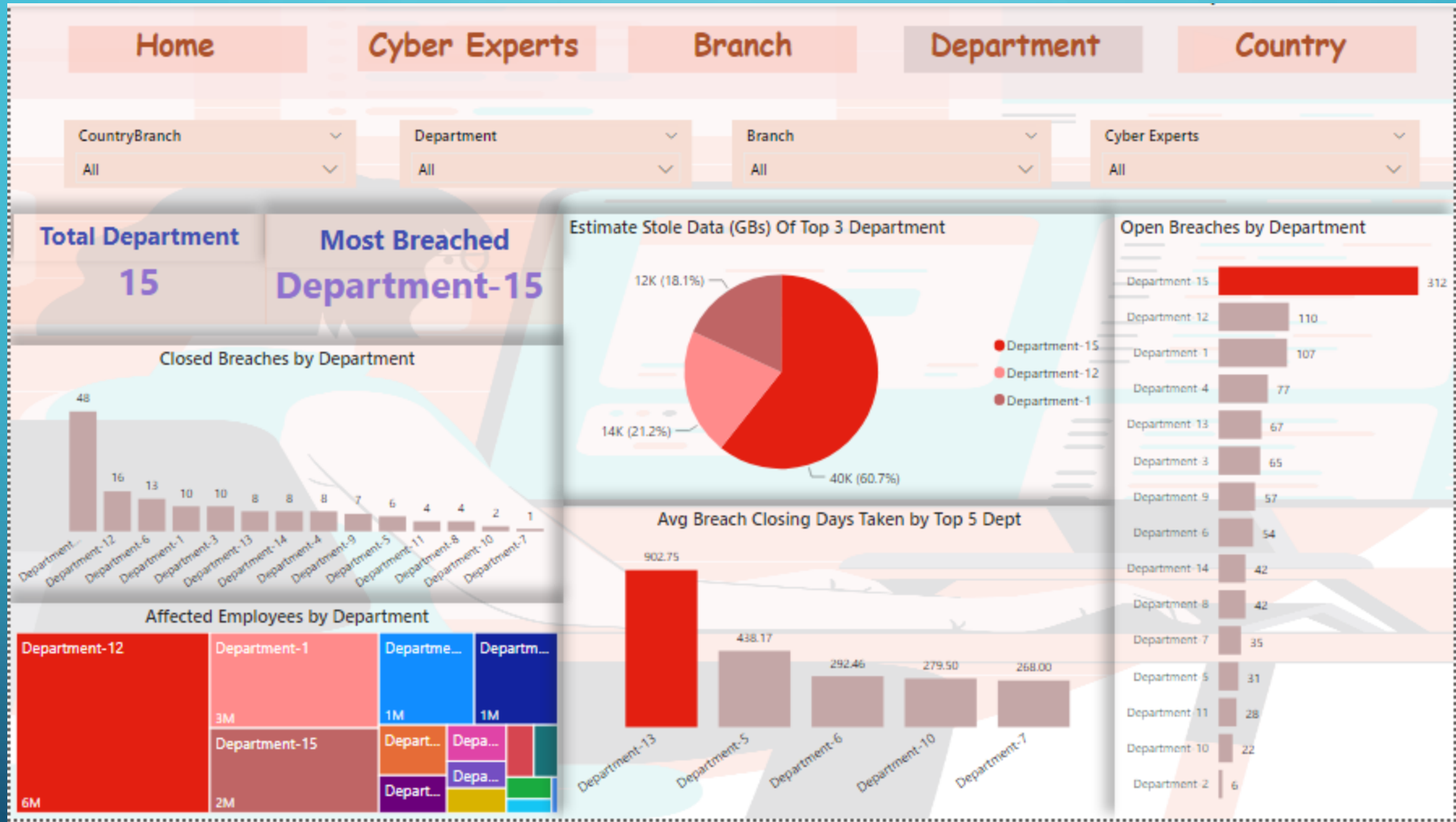
Affected Employees  
16M  
81.96%

Most Breached Country  
Canada

Secured Employees  
3M  
18.04%












# VISUAL ANALYTICS AND INSIGHTS

- ✓ There are total 1055 breaches detected from which 145 are closed and 910 breaches are still open which is 86% of total breaches.
- ✓ Average Time taken to Close a Breach is 243 days because of which 138K GBs of data was stolen.
- ✓ Out of 19M employees, 16M employees are victim of breaches which is 84% of total employees.

- 
- ✓ From all 11 cyber experts, top 4 experts have detected 51% of total breaches.
  - ✓ From all 5 branches, Branch-2 is most impacted branch with 349 total breaches out of which 314 breaches are open, 64K GBs of Stolen Data and 9M affected employees.
  - ✓ From 15 departments, Department-15 is most breached department with 312 breaches out of which 264 breaches are open and 40K GBs of stolen data. Also, Department-15 has closed 48 breaches which highest among all departments.
- 

- 
- The background is a solid blue gradient. In the corners, there are decorative white line art elements resembling circuit boards or neural networks, with lines and small circles connecting them.
- ✓ From 25 Countries, Canada is most breached country with 55 breaches and 7.4K GBs of Stolen data.
  - ✓ Canada, India and United States have most open breaches, each have 49 open breaches.
  - ✓ Sweden and Norway has highest Average Time taken to Close Breach of 665 Days.
  - ✓ United States has highest number of employees(5M) which are affected from various types of breaches.

The background is a blue gradient with decorative white circuit-like lines in the corners. The lines consist of straight segments and small circles, resembling a stylized electronic circuit.

THANK YOU  
PRESENTATION BY : AMIT SANGWAN