الأكاديمية السعودية الرقمية
SAUDI DIGITAL ACADEMY

وزارة الاتصالات وتقنية المعلومات
MINISTRY OF COMMUNICATIONS AND INFORMATION TECHNOLOGY

CODING DOJO

CYBER CRIMES IN SAUDI ARABIA

Group Members :

- Raghad Almeataz
- Ahmed Altowairqi
- Mohammed Albesher
- Abdullah Albutih
- Abdulrahman Aljubaylan

Business Understanding

Overview

- Cyber crime cases are drastically increasing in Saudi Arabia. At present, Saudi Arabia is one of the top 10 countries in significant cyber attacks.

- In addition, according to specialists in information security, the financial loss estimated by Saudi companies as a result of piracy and electronic sabotage ranges from SR 300,000 to more than SR1 million for each case, with bank losses, estimated at more than a billion US dollars.

Overview

- The 2030 VISION OF SAUDI ARABIA seeks to Increase the Country Income that is not based on Oil that include Information Technology Sector where decreasing the Internet Security Threats to save the lost Money and to use the lost money in other beneficial areas such as investments and Organizations Development and other.

Banks Financial Losses	4.5B
Companies Financial Losses per Case	300K - 1.2M
Number of Victims in 2016	3.6M

All Financial Losses are in SAR Currency - SAUDI RIYALS

Project Objective

- measure the awareness of Saudi people about the danger of cyber crimes.
- find out the major factors for becoming a potential victim.
- give Solutions for the Main faced issue .
- measure the Risk factors and decrease it .

Stakeholders

The Target Audience

- Cyber security companies
- Saudi Federation for Cybersecurity, Programming and Drones Institute (SAFCSP)
- Businesses (organizations) that want to measure the awareness of their employees
- Any company that is working remotely, needs to check security applications of their own
- employees to ensure consistent work

Data Collection & Preparation

Dataset Overview

- It is a survey conducted using Google forms on Saudis whose age is above than 18. It consists of 64 questions, which indicates the number of columns and 1230 number of responses which indicates the number of rows.

64 Questions

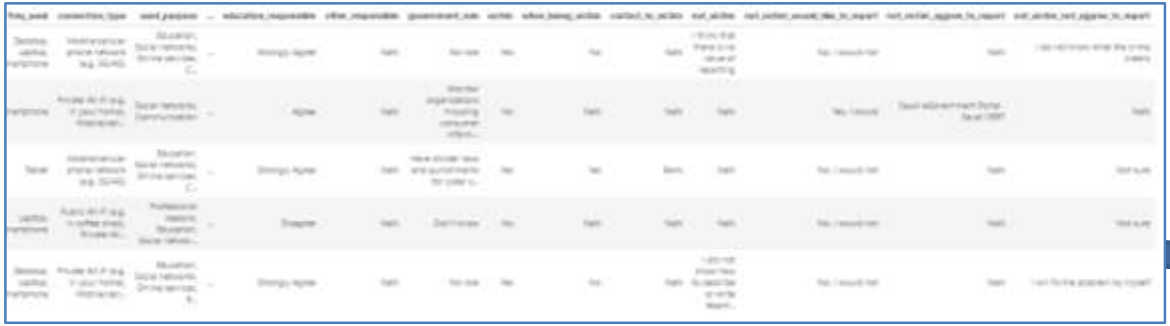
1230 Responses

Rows : 1230 | Columns : 64

Data Preprocessing

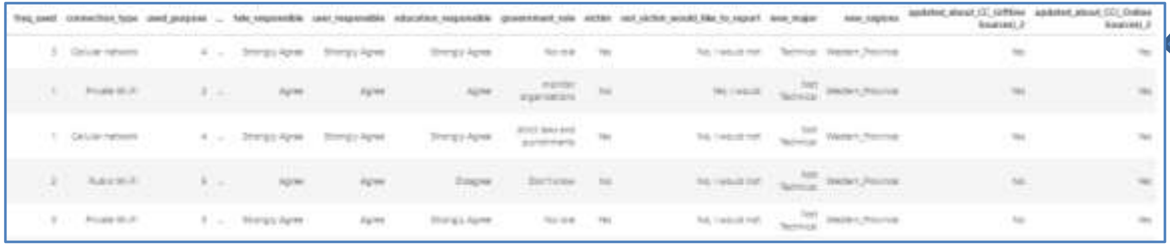
The applied Approaches :

- Filling with zeros
- Filling with Values
- Dropping The Null Values
- Filing with Forward Filling
- Filling with Backward Filling



freq_used	connection_type	used_purpose	...	education_responsible	other_responsible	government_role	victim	when_being_victim	contact_to_victim	not_victim	not_victim_would_like_to_report	not_victim_approve_to_report	not_victim_not_approve_to_report
Desktop, Laptop, Smartphone	Mobile/cellular phone network (e.g. 3G/4G)	Education, Social networks, Online services, C...	...	Strongly Agree	NaN	No one	Yes	No	NaN	I think that there is no value of reporting	No, I would not	NaN	I do not know what the crime means
Smartphone	Private Wi-Fi (e.g. in your home, Workplace)...	Social networks, Communication	...	Agree	NaN	Monitor organizations, Provoking consumer interest...	No	NaN	NaN	NaN	Yes, I would	Saudi eGovernment Portal... Saudi CERT	NaN
Tablet	Mobile/cellular phone network (e.g. 3G/4G)	Education, Social networks, Online services, C...	...	Strongly Agree	NaN	Have stricter laws and punishments for cyber c...	Yes	Yes	Bank	NaN	No, I would not	NaN	Not sure
Laptop, Smartphone	Public Wi-Fi (e.g. in coffee shop), Private Wi...	Professional reasons, Education, Social network...	...	Disagree	NaN	Don't know	No	NaN	NaN	NaN	No, I would not	NaN	Not sure
Desktop, Laptop, Smartphone	Private Wi-Fi (e.g. in your home, Workplace)...	Education, Social networks, Online services, E...	...	Strongly Agree	NaN	No one	Yes	No	NaN	I do not know how to describe or write report...	No, I would not	NaN	I will fix the problem by myself

After conducting the different approaches, we continue with the best fit which is the **Dropping The Null values.**



freq_used	connection_type	used_purpose	...	tele_responsible	user_responsible	education_responsible	government_role	victim	not_victim_would_like_to_report	ww_major	ww_region	updated_about_CC_(offline Sources)_2	updated_about_CC_(Online Sources)_2
3	Cellular network	4	...	Strongly Agree	Strongly Agree	Strongly Agree	No one	Yes	No, I would not	Technical	Western_Province	Yes	Yes
1	Private Wi-Fi	2	...	Agree	Agree	Agree	monitor organizations	No	Yes, I would	Non Technical	Western_Province	Yes	Yes
1	Cellular network	4	...	Strongly Agree	Strongly Agree	Strongly Agree	strict laws and punishments	Yes	No, I would not	Non Technical	Western_Province	Yes	Yes
2	Public Wi-Fi	5	...	Agree	Agree	Disagree	Don't know	No	No, I would not	Non Technical	Western_Province	No	Yes
3	Private Wi-Fi	3	...	Strongly Agree	Agree	Strongly Agree	No one	Yes	No, I would not	Non Technical	Western_Province	No	Yes

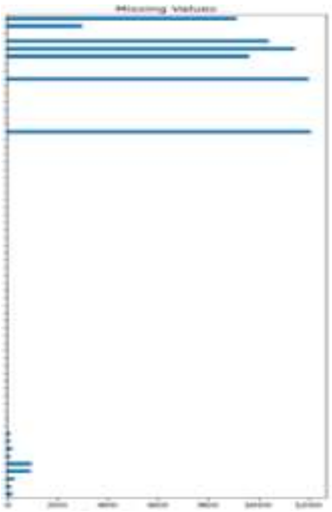
Data Preprocessing

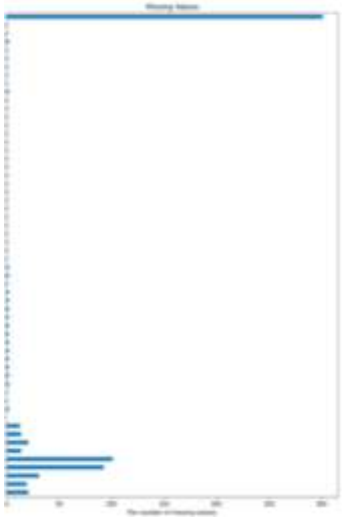
we dropped the columns that with 60% missing values and more , because they do not have enough data

The Original Dataset

60*1231 => 73860 / 100 => 739

Dropping Null Values

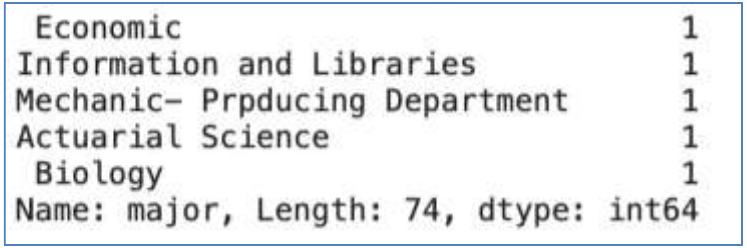




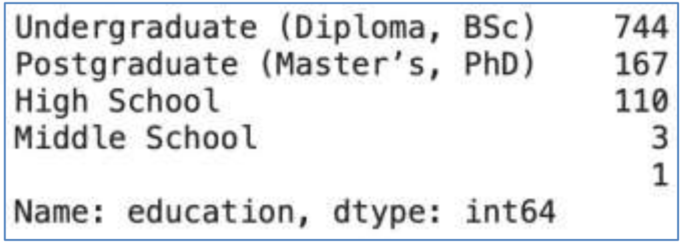
Missing Values
−0.04
−0.02
0.00
0.02
0.04
The number of missing values

Data Preprocessing

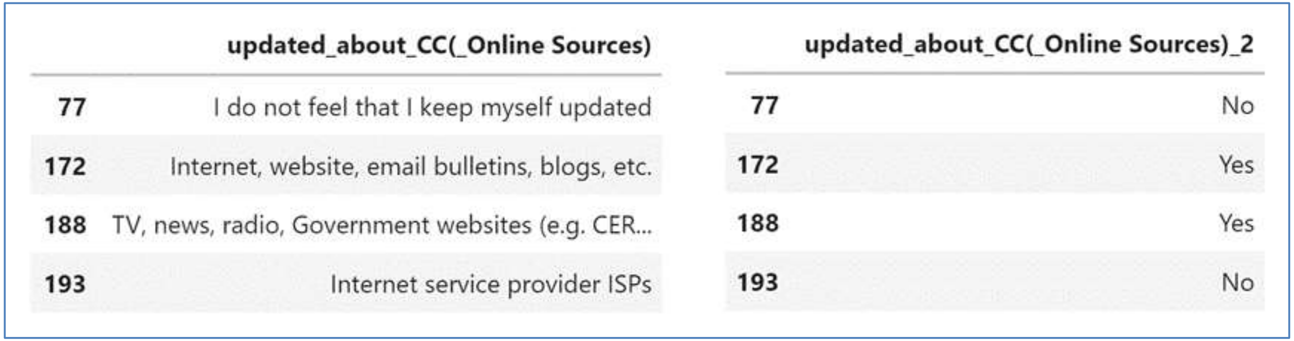
Trim white spaces, imputing empty values and extracting new columns out of existing ones.



```
 Economic                                1
Information and Libraries                1
Mechanic- Prpducing Department           1
Actuarial Science                        1
 Biology                                 1
Name: major, Length: 74, dtype: int64
```



```
Undergraduate (Diploma, BSc)    744
Postgraduate (Master’s, PhD)    167
High School                     110
Middle School                     3
                                  1
Name: education, dtype: int64
```



	updated_about_CC(_Online Sources)
77	I do not feel that I keep myself updated
172	Internet, website, email bulletins, blogs, etc.
188	TV, news, radio, Government websites (e.g. CER...
193	Internet service provider ISPs

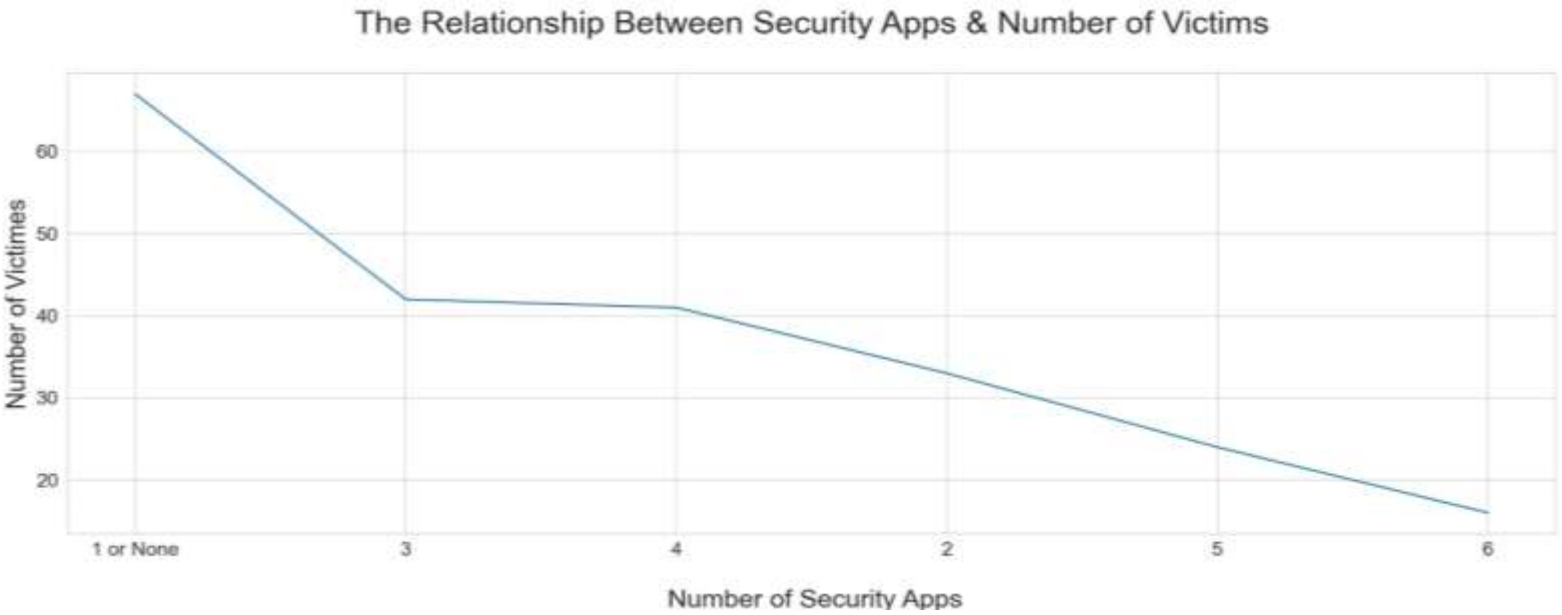
	updated_about_CC(_Online Sources)_2
77	No
172	Yes
188	Yes
193	No

Exploratory Data Analysis

EDA (Factors)

❖ **Insight** :

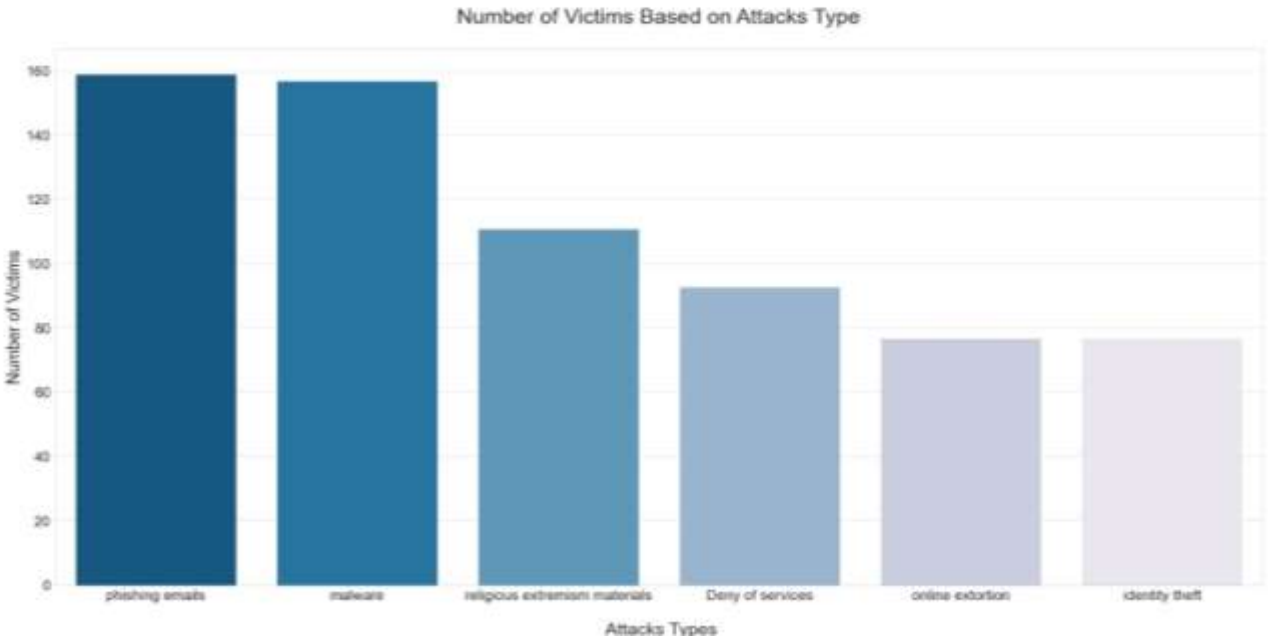
- The less security apps the more potential of becoming victim.



EDA (Factors)

❖ **Insight** :

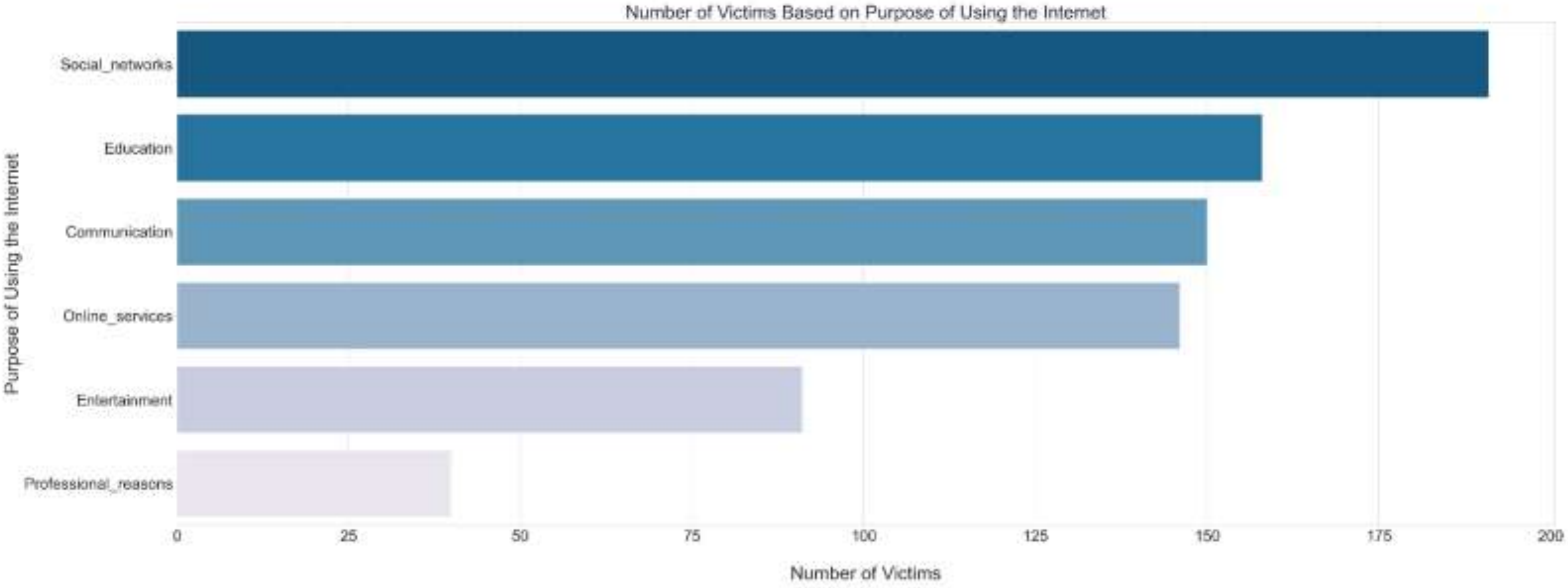
- Which type of attacks has the most effect on becoming a victim?



EDA (Factors)

❖ **Insight** :

- Which purpose has the most effect on becoming a victim?



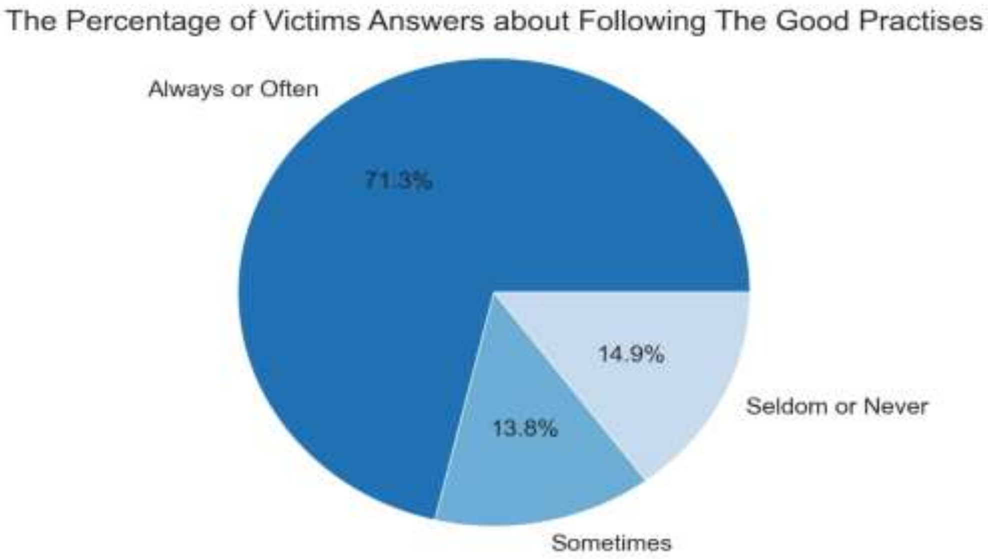
EDA (Factors)

❖ **Insight** :

- The more good practices you follow the less potential of becoming a victim

Good Practices:

1. Careful clicking on links
2. Install software updates
3. Change pass frequently
4. Terms conditions
5. Privacy settings
6. Check legitimacy

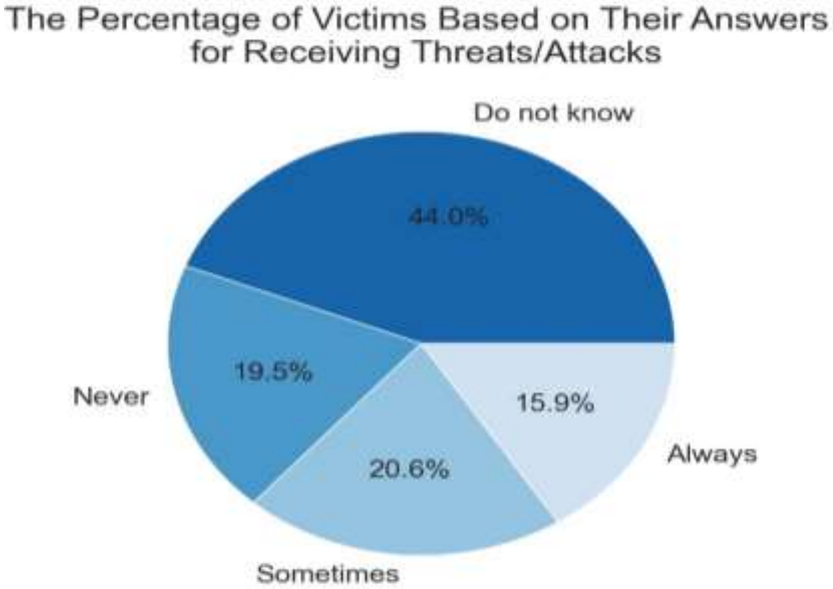


The Percentage of Victims Answers about Following The Good Practises

EDA(Awareness)

❖ **Insight** :

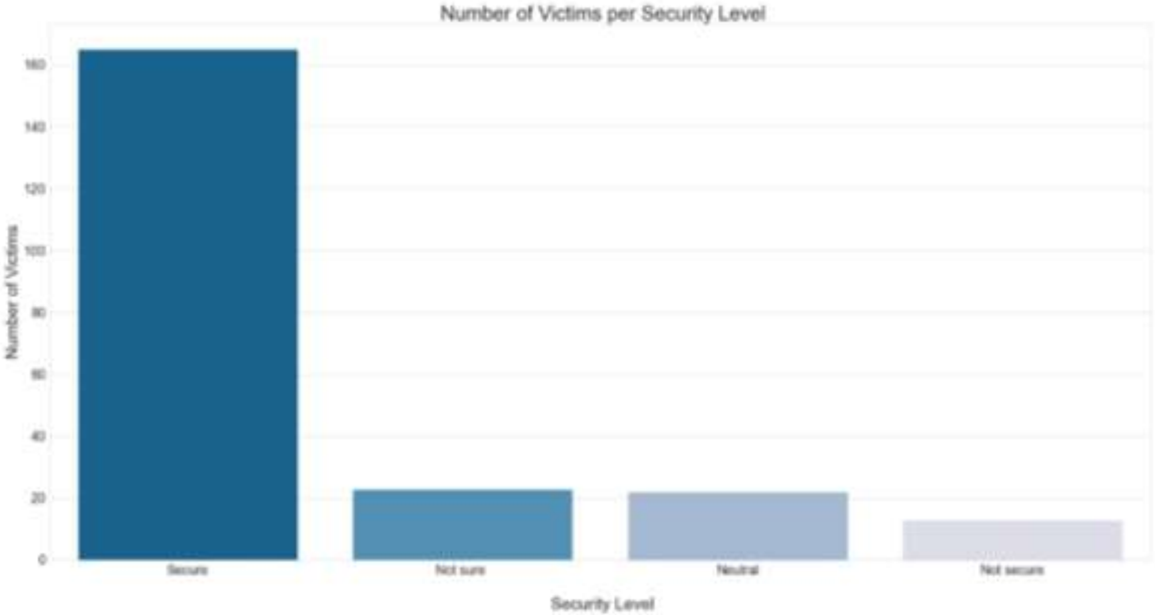
- People always receive threats/attacks



EDA(Awareness)

❖ **Insight** :

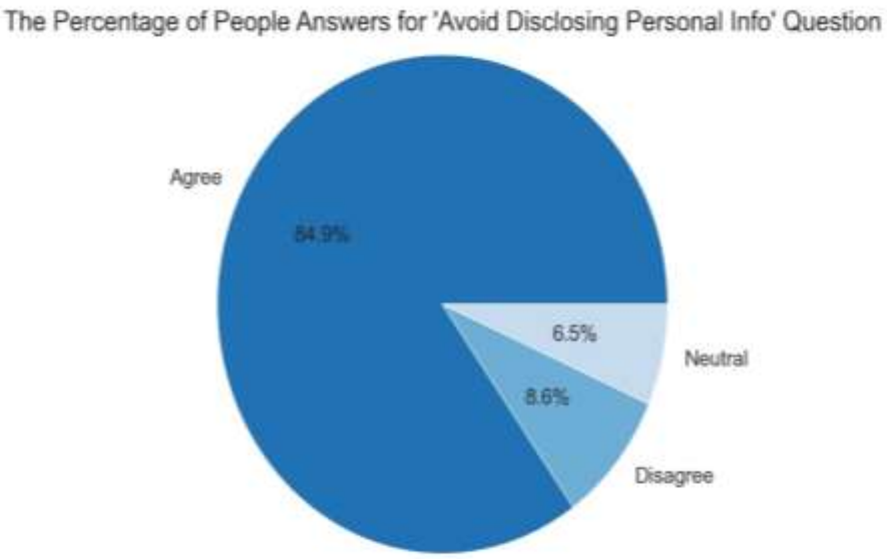
- The more security level the less potential of becoming victim

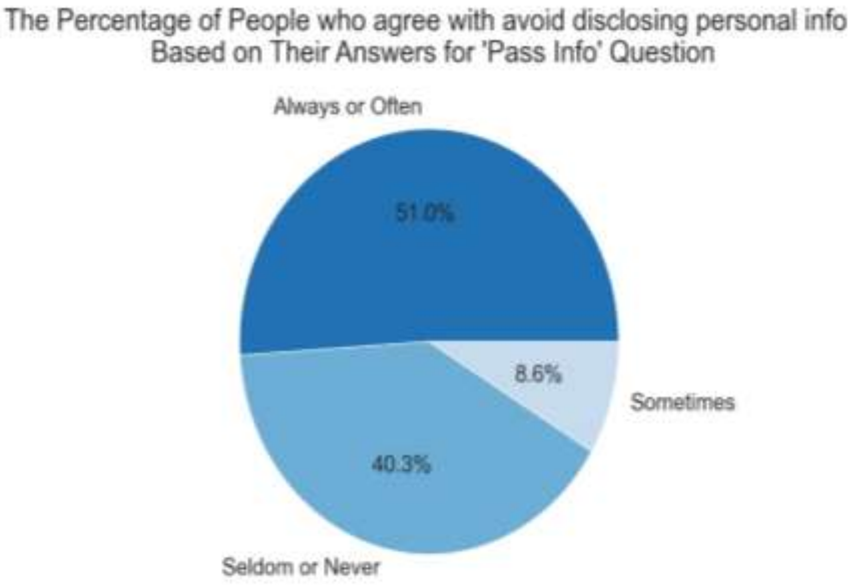


EDA(Awareness)

❖ **Insight** :

- Most people agree with avoiding disclose personal information.



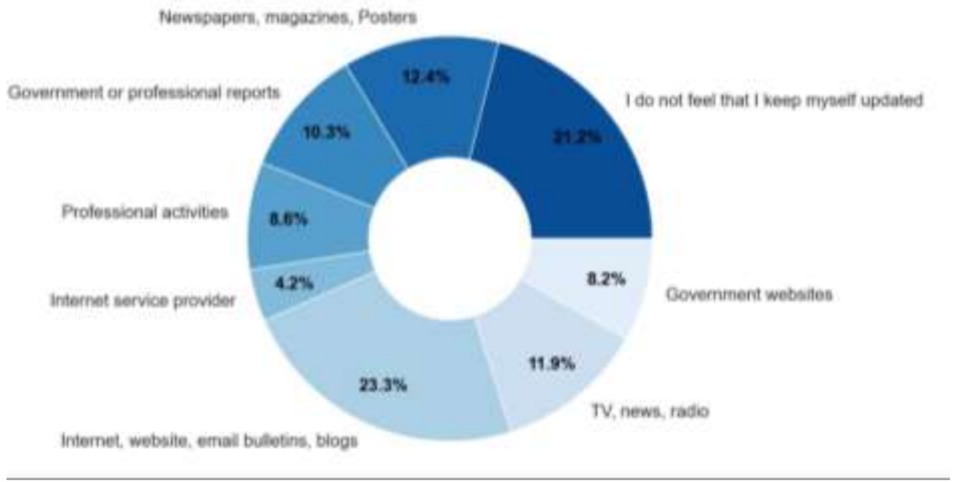


EDA(Awareness)

❖ **Insight：**

- What are the top channels that people use to raise their awareness on cyber crimes?

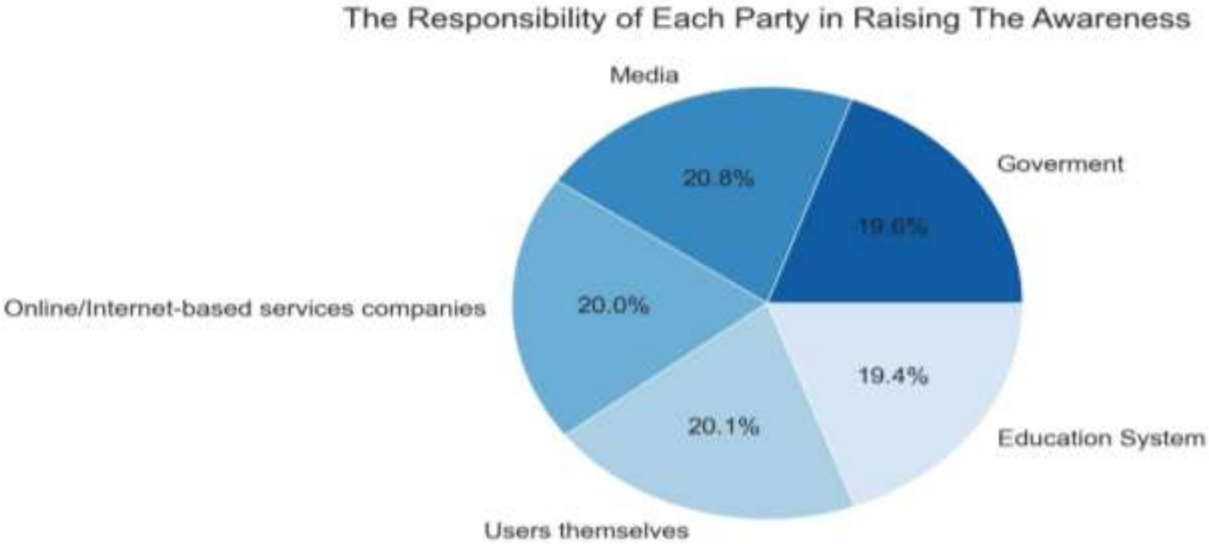
The percentage of channels that people use to raise their awareness about cyber crimes



EDA(Awareness)

❖ **Insight：**

- Which party is more responsible for raising the awareness of cyber crimes?



Modeling

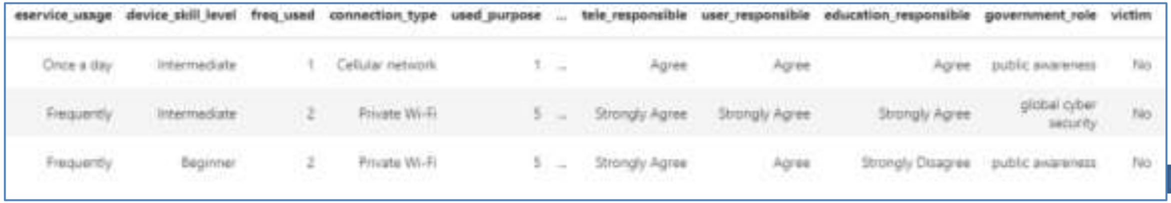
Modeling Overview

The Dataset Name	Cyber crimes
The Target Label	Have you been a victim of a cyber crime ?
The Machine Learning Type	Supervised Learning
The Modeling Category Type	Classification
The Programming Language	Python
The Main Library	Scikit-learn

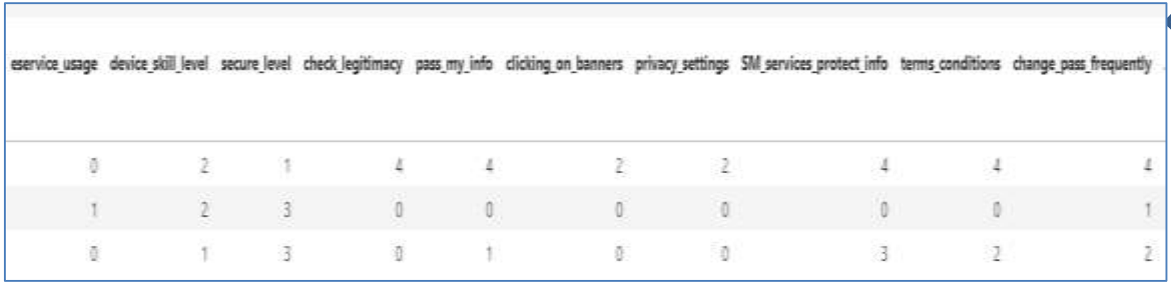
Feature Engineering

The applied Mechanisms :

- Mapping
- Label encoding
- One Hot encoding



eservice_usage	device_skill_level	freq_used	connection_type	used_purpose	...	tele_responsible	user_responsible	education_responsible	government_role	victim
Once a day	Intermediate	1	Cellular network	1	...	Agree	Agree	Agree	public awareness	No
Frequently	Intermediate	2	Private Wi-Fi	5	...	Strongly Agree	Strongly Agree	Strongly Agree	global cyber security	No
Frequently	Beginner	2	Private Wi-Fi	5	...	Strongly Agree	Agree	Strongly Disagree	public awareness	No



eservice_usage	device_skill_level	secure_level	check_legitimacy	pass_my_info	clicking_on_banners	privacy_settings	SM_services_protect_info	terms_conditions	change_pass_frequently
0	2	1	4	4	2	2	4	4	4
1	2	3	0	0	0	0	0	0	1
0	1	3	0	1	0	0	3	2	2

After conducting the different mechanisms and try them all, we continue with the best fit which is mix between **the Label encoding and one hot encoding.**

Challenges & Solutions

The Challenges :

- **The label Imbalance**.
- **The Dataset Complexity**.
- **beating the baseline model**.

The Solution :

After Conducting many methods to fix the faced issues, we found out that the best solution is to remove around 200 records from the dataset because going with this method would not just resolve the missing values issue but also would resolve the bias in the dataset and makes it more suitable and equivalent. Beating the baseline model in the start was a rough process but after removing the specified records and decreasing the bias in the dataset now the data classes gap minimized which makes more sense.

Models

The Models Results :

The Baseline	**0.7707317073170732**
The Logistic Regression	0.824390243902439
The Decision Tree	0.824390243902439
The Random Forest	**0.8146341463414634**
The SVC	0.824390243902439

Models

Why did we chose this specific Model ?

The Random Forest model characterized by Low risk of overfitting and runs efficiently on a large dataset.

We chose the Random Forest model because it’s the best model to represent Great accuracy with Strong reliable almost parallel features that all the entered features contributing to the prediction.

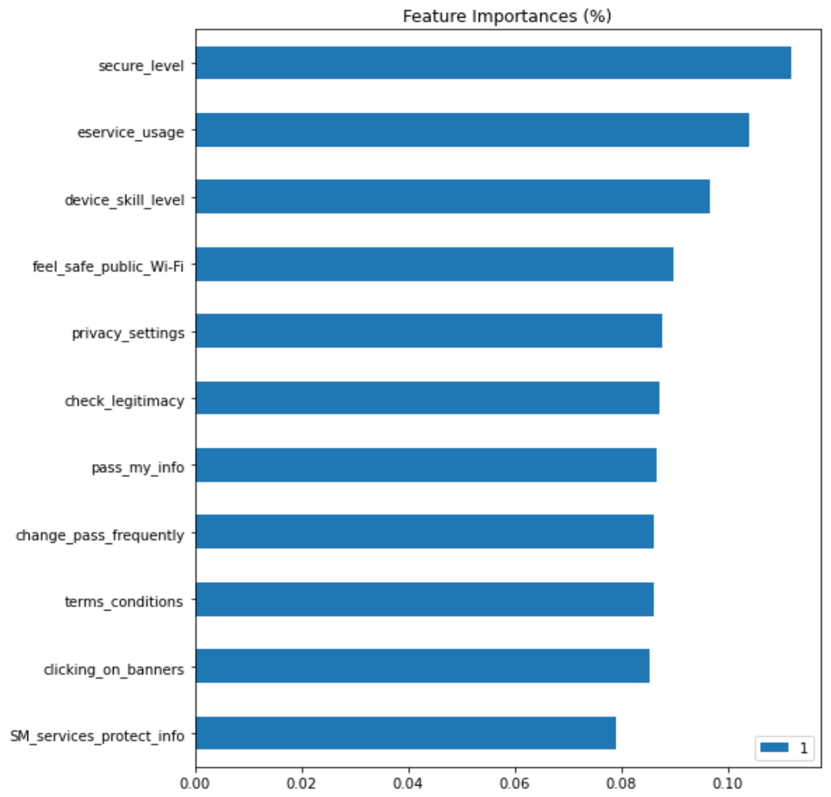


Tableau Dashboard

[The Dashboard]

Conclusion

We proposed a cybercrime Dataset that:

- ❗ Filled with Messy Data
- ❗ Filled with biased Data
- ❗ Lack of Specification

After we finished the project , we could :

- ✓ Measure The awareness of people
- ✓ Get The Main factors that Couse cybercrimes
- ✓ Determine if a person is going to be a cybercrime victim or not

Future Work

In the near future :

- Add the project into Different blogs.
- We Share The Results with the Local Community
- We seek to continue to develop the project.
- Publish The work and add more Data.

Acknowledgement

Mr. Mikio Harman for his time, effort and dedication to helping us. We learned a lot from him and enjoyed working with him.

Saudi Digital Academy for supporting us and provide us with this wonderful Data Science Bootcamp that gave us the Strong Background we need so we can take our Careers to the Next level.

Coding dojo for supporting us and provide us with all the Resources we need to Increase our knowledge and work practically with different creative new projects.

THANK YOU

Any Questions ?