Table Of Contents

INTRODUCTION

This project focuses on data exploration of the fire incidents that took place in Toronto,Ontario from the years 2011-2016. The dataset gives the detailed information about the places in Toronto where the fire incidents have been recorded in the city. The location of the places is specified by the latitude and longitude of that place. The dataset consists of many different columns which gives the details of the fire incidents such as :
1) Area of the origin which gives the place at which the fire started.
2) Ignition source - the type of object from which the fire was ignited.
3) Impact of evacuation which includes how they were evacuated.
4) The method by which the fire got controlled.
5) Possible cause of the fire and so on.

The user interactive visualisations help to analyse different parameters such as number of accidents that took place on a particular day according to the hour of the day for example the count of no of fire incidents took place on sunday at different hours i.e 21 hour, 23 hour etc. This visualization helps to find out the risky day at which the number of fire incidents is the highest and it can also explain that at which hour the no of incidents is the maximum compared to other hours. With the help of this we could find the most risky day and the most risky hour of the fire incidents and people can take certain measures to avoid the fire incidents in their houses. Such as do not leave the premises unattended, do not leave any dangerous thing which could light up the fire in the house, be more careful at that particular time etc. Second visualisation explains the importance of the fire alarm system presence in the house according to the dollar loss in the house. This visualisation gives dollar losses that took place according to the presence of the fire alarm. The dollar loss is being accumulated into ranges such as $0-$500,$500-$5000 etc. It gives the estimated dollar loss range when the fire alarm is present and it also gives the estimated dollar loss range when the fire alarm is not present. With the help of this visualization we could say that if the fire alarm is present how much could be the dollar loss and vice versa. This visualization explains the importance of the fire alarm as it gives the count of the dollar loss that lies in that range. Third visualisation gives the relationship between the three variables which is presence of fire alarm , working of the fire alarm and the method of the evacuation executed. In this visualization it shows the percentage of the relation of the three variables. With the help of this visualization it shows the importance of the fire alarm system presence of the fire alarm and working of the fire alarm. This visualization can analyse how many % of the people evacuated with the fire alarm presence and the working or how many % of the people evacuated with the fire alarm not present and not working and how the evacuation executed etc.

The intended audience for this project are the general public. They can use this data and learn from the visualizations how the fire alarm system’s importance is seen during cases of fire. They can clearly see from the visualizations how the impact is seen on dollar loss and evacuation of the people. This will help them to focus more and pay more attention to any defaults if present in

their fire alarm systems. The project will also help the general public to see which hours of the day incurs the most number of incidents and hence they can be much more careful during those times.

DESIGN

After the data exploration of the data many different questions came into the mind. This helps to prepare different visuation to answer different questions. First of all I made many visualizations which contain graphs like line charts , bar charts,heatmaps , sunburst, leaflet map,animated line charts, etc. This was the brainstorm which contains all the ideas and all the visualization I could make according to the data. Then after that I selected the relevant visualizations according to the question that I have to answer and the sophisticated visualization which is quite extensive, informative, easy to read and also gives a story telling.
After that I picked up the best visualization to answer the particular question which contains the user interaction data and is most informative from all the visualization for the same question that we are answering.

The five design methodology is attached in Appendix.

Explanation of the five design sheet methodology :

Sheet 1
This sheet contains all the ideas that came across to solve the different types of questions with different visualizations. This sheet contains different charts showing different relationships between variables which give the answers of the many different questions.
It contains many charts such as heatmap,barcharts,sunburst, Heat map,cluster leaflet map etc.

Sheet 2
This sheet contains the initial design 1 which contains the visualization of answering my question i.e stacked barchart , heatmap and a scatter chart. The heatmap is the main focus here which answers the question of displaying which day sees the most number of incidents and in which hour of the day. The rest of the visualizations were not able to answer the question appropriately and hence were not selected. These rejected visualizations were first tried in R but because they were less informative they were not selected.

Sheet 3

This sheet contains the initial design 2 which contains the visualization of answering my question i.e streamgraph, linechart , sunburst. The sunburst is the main focus for this sheet as it clearly displays all the required information and answers the question as well. It is an interactive visualization which lets the user to hover from inside out to get the percentage values. The rest of the visualizations including the linechart and the streamgraph were rejected as they were not informative and did not convey the message correctly. Hence,they do not help to deduce information appropriately. So we reject these visualizations.

Sheet 4
This sheet contains the initial design 3 which contains the visualization of answering my question i.e animated line chart , stacked bar chart and a streamgraph with hovering additionality. The main focus for this was hence the animated line chart which displays the relevant information of dollar loss based on the absence/presence of the fire alarm system. The other visualizations did not support and portray the data to the best extent and hence they have not been used. In animated graphs we can select the presence of the fire alarm with the help of the selection the graph changes. These graphs give detailed information on why these visualizations are being accepted.

Sheet 5
In this sheet my accepted visualizations are being accumulated in one shiny dashboard which contains different tabs with different visualizations. This contains user interactive data as we select the different tabs then the visualization changes. User interaction also includes the hovering in the charts to get detailed information. This sheet hence contains the main focuses of the above sheets including the sunburst, heatmap and animated line chart.

IMPLEMENTATION

For the implementation of the visualisations, certain manipulations to the data were made in the exploration project. To implement the sunburst graph, I have created a dataframe according to the needs of the graph. This sunburst shows how the Presence/Absence of the Fire Alarm System and how the functioning of it affects the evacuation of the people. Hence describing how adversely the people were affected by the fire. The three categories/columns that have been used are the **Fire Alarm System Presence, Fire Alarm System Functioning and the Evacuation of People.** The NA values in these columns were removed during the exploration project. Then after removing the NA values, a groupby function has been used to get the count of the incidents based on the three categories. Using this data a Sunburst has been created. By hovering from inside out in this Sunburst, we can see a percentage of the different combinations of these three categories. From this visualization we can infer about the percentage evacuation

of people. This has been implemented in the R shiny Dashboard under the tab named **Sunburst.**

The visualization looks as shown below -

Figure 1 - Sunburst to show relationship between Fire Alarm System Presence, Fire Alarm System Functioning and the Evacuation of People

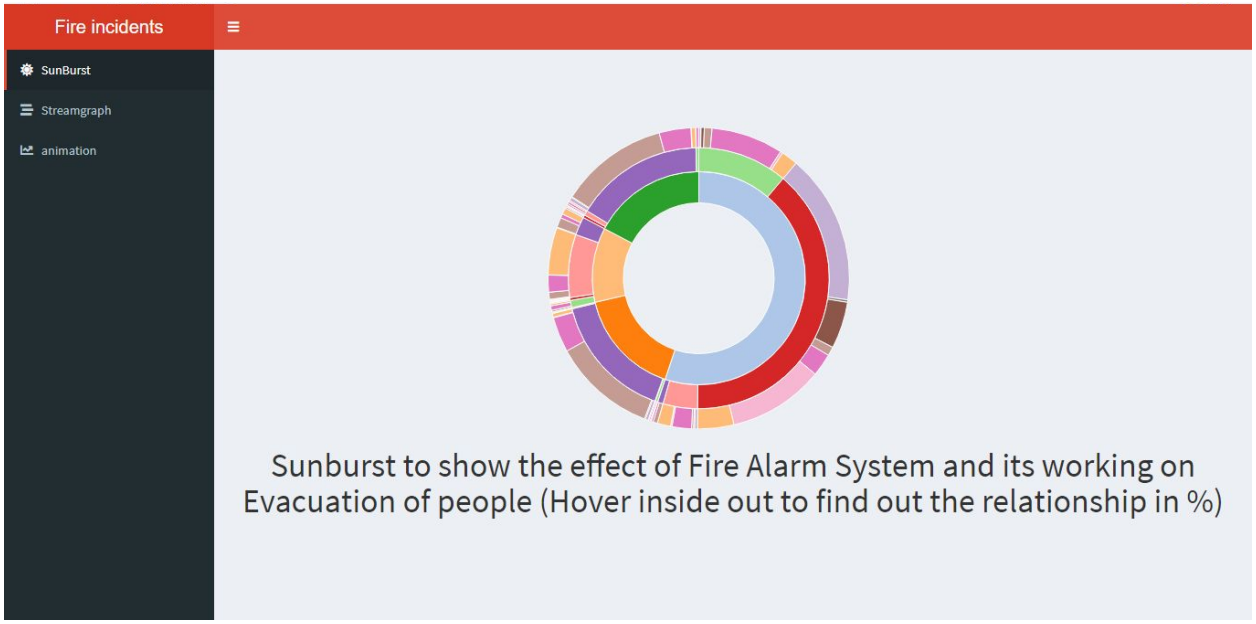
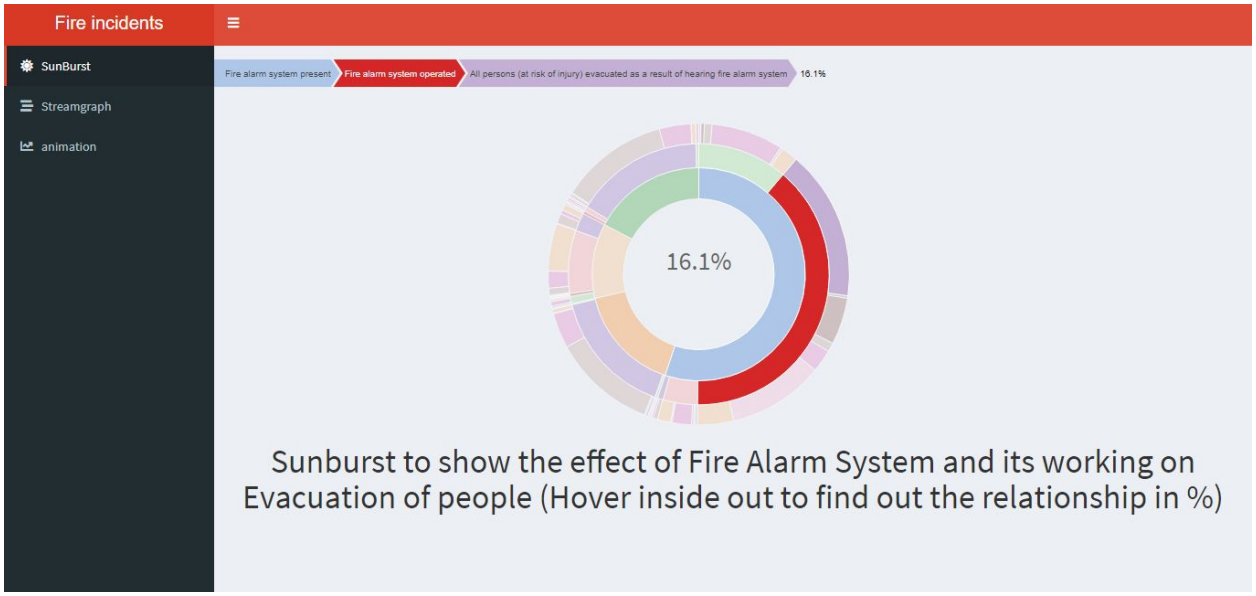


Figure 2 - Hovering in the Sunburst



The next visualisation is the Heat Map which depicts the hour wise count of the incidents based on each day of the week. For the implementation of this visualization, I have created a dataframe that contains the day wise and hour wise count of the fire accidents in Toronto. The groupby function has been used in R on the columns **HourWise, DayWise** to finally get the count for each row(Day-Hour combination). This data frame hence contains the data in the required format. Using this dataframe a Heat Map has been created in shiny R under the tab **Heat Map**. On hovering above the HeatMap we can see the exact day, hour and the value. This helps the user to determine the exact counts and check which day or hour sees the maximum count of incidents. There is also an option to zoom in the heat map.

The visualization looks as shown below -

Figure 3 - Heat Map to show Hour wise and Day Wise incidents count

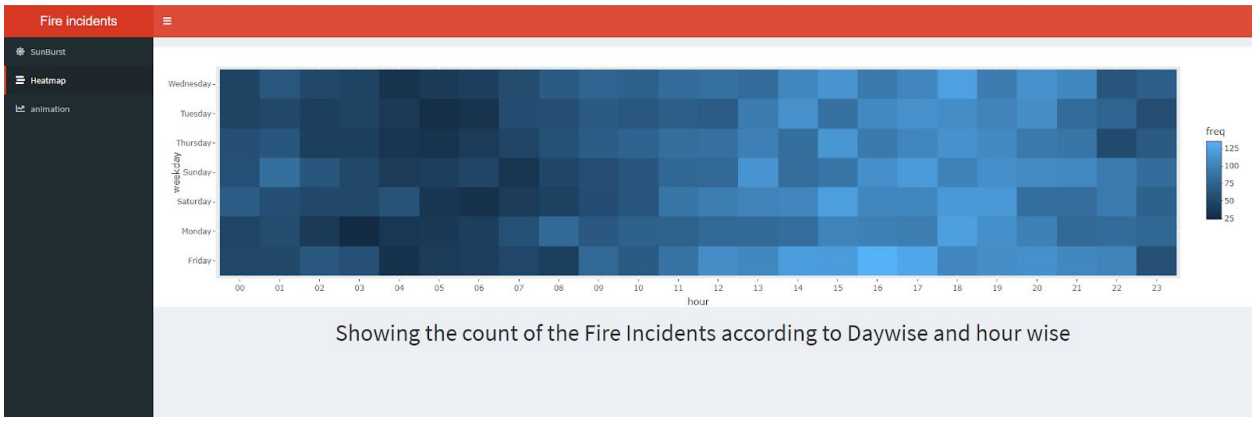
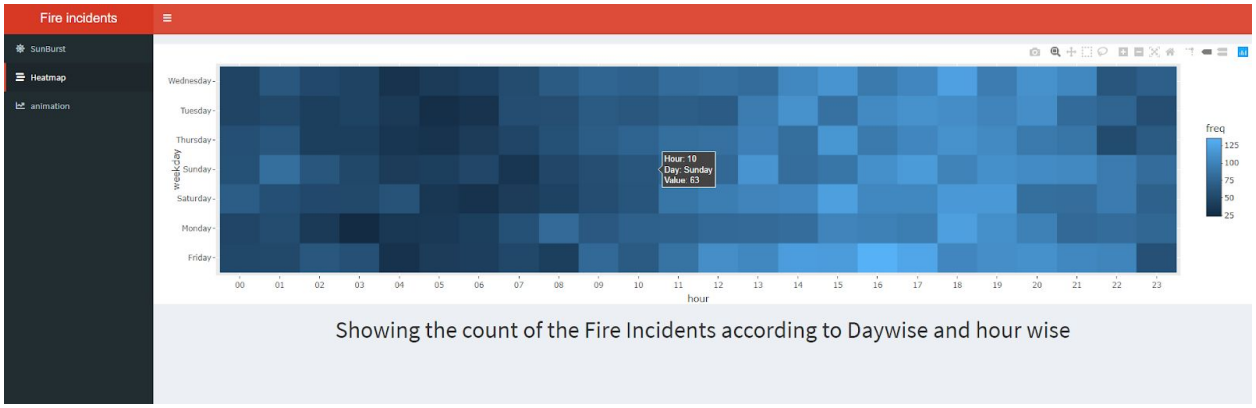


Figure 4- Selection of a day highlights the Heat map

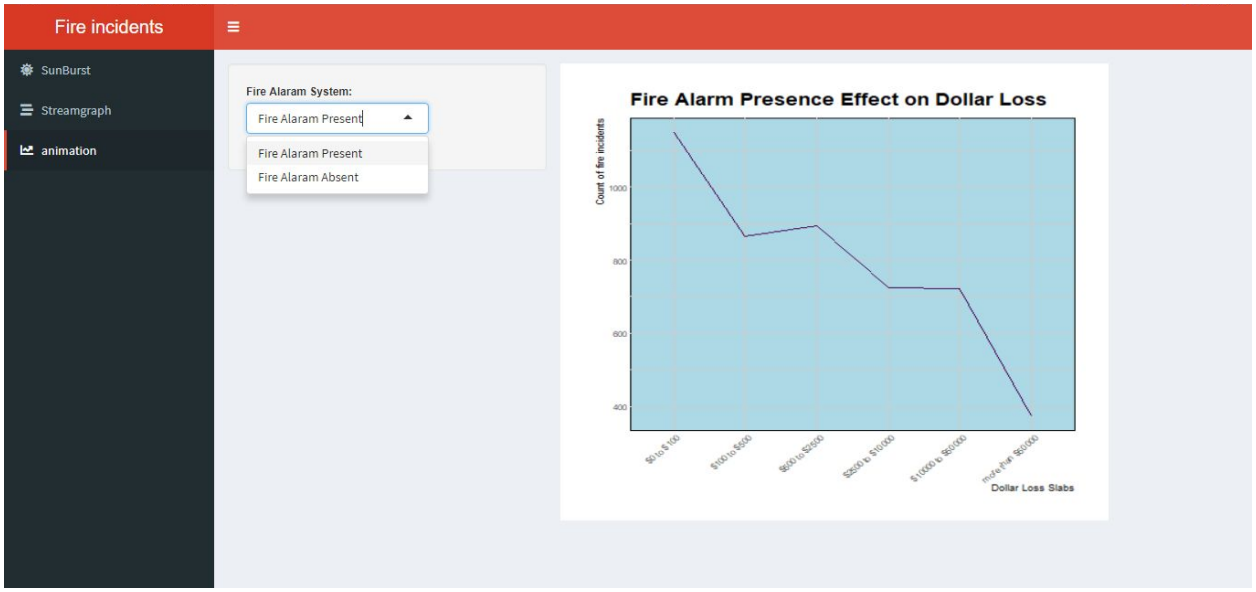


The final visualization contains an animated line chart. This line chart shows how the presence and absence of a Fire Alarm System affects the dollar loss. This animated line chart has been implemented by using the columns **dolar loss.** As the values are discrete hence slabs of dollar loss have been created. Then a new dataframe has been constructed by counting the occurrence of each slab giving the incidents in each dollar loss slab. Now an animated line chart

has been used giving it the corresponding x and y axis. Rendering of these animated line charts(one for presence and one for absence of fire alarm system) happen initially in the beginning only and then the corresponding saved values are used in the server part of the shiny R application.

The visualization looks as shown below -

Figure 5 - Animated Line chart



The libraries used for implementation are -

library(shinydashboard)
library(shiny)
library(tidyr)
library(dplyr)
library(plotly)
library(viridis)
library(quantmod)
library(ggplot2)
library(gganimate)
library(sunburstR)
library(gifski)

USER GUIDE

The User can run the shiny app and will be displayed with a R shiny Dashboard. This dashboard contains visualizations that will guide the user to visualise different aspects related to fire incidents in toronto. There are three tabs displayed on the right hand side of the dashboard.

The user can click on the first tab named Sunburst. By clicking on this tab the user is displayed with a Sunburst Graph which shows how the Fire Alarm System's Presence/Absence and its functioning has an affect on the evacuation of people. The user can hover/ move his or her mouse from inside out of the Sunburst to see the relevant percentages.

The user can now click on the second tab named as Heat Map. On clicking this tab the user will be displayed with a Heat Map that shows the fire incidents based on the day of the week and the hour wise count. The user can hover over the heat map to get exact counts of the incidents for a particular day and for a particular hour of the day. The legend on the right also shows the range in the colour based on the count of the incidents. The hours appear on the x axis and the counts appear on the y axis. Hence the user can see the counts and find the most common hour/time when the fire incidents take place.

The user can now select the third tab named as Animation to see the relation between the presence/absence of the fire alarm system with the dollar loss incurred. The user can select from the drop down of whether the Fire Alarm System was Present or whether the Fire Alarm System was Absent. Based on the selection of the user the animated line chart changes.

CONCLUSION

After all the visualization and answering the question above, the interpretation we got is that the most number of fire incidents that took place in Toronto is on Thursday,Friday,Saturday and Sunday from 13 hours to 21 hours. These counts are hence very high as compared to the other days and other hours of the day.
Another conclusion came across was that the relationship between the three variables(fire alarm present,operation of the fire alarm and the evacuation execution type).With the help of sunburst visualization it depicts the relationship between the three variables in terms of percentage. This visualization tells us that when there was a fire alarm system present and when it was working the most number of people were evacuated from the scene of the incident. And when the fire alarm system was not present or when it was not working, a few percentage of people were evacuated.

From the animated line chart we can see that when the fire alarm system is present then the dollar loss is seeing a decreasing trend where a huge dollar loss is seen the minimum number of times. Similarly when the fire alarm system is absent then the animated line chart sees an increasing trend indicating that a huge dollar loss is seen the maximum number of times. Hence this helps us to conclude that a fire alarm system presence will have a high chance of reducing the dollar loss.

From the above Project of the fire incidents in toronto I have learned how to read the data in an efficient manner and learned to explore the data to retrieve answers from the data. With the help of this project I got to learn many data visualization techniques which helped me to visualize the data and answer different types of questions. I have also learned to create new and different types of visualization which includes interactive methods such as sunburst ,heatmap, shiny R app server etc.Hence this project helped to to make insights of the data and to make different visualization.

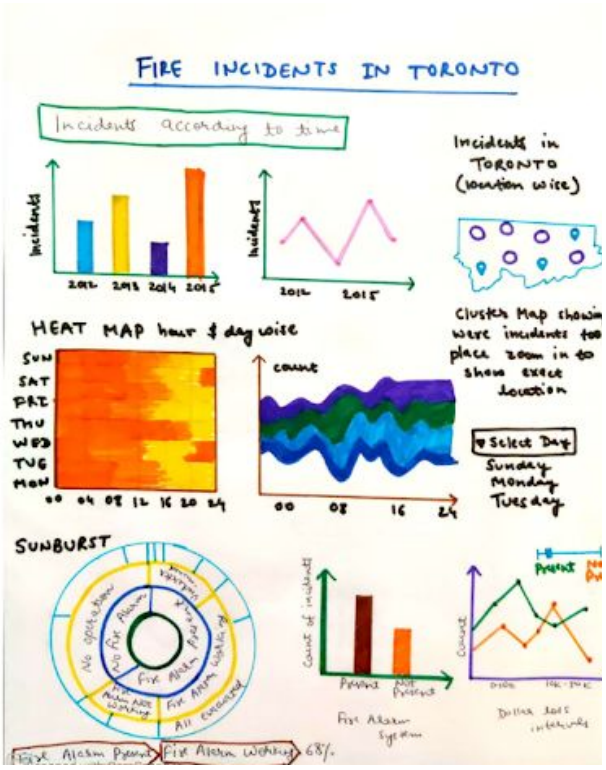
For further improvement of the results I would have added more visualizations to answer deeper questions like why the particular hour of the day sees the most number of incidents and so on.

APPENDIX

FIVE DESIGN SHEET

SHEET 1

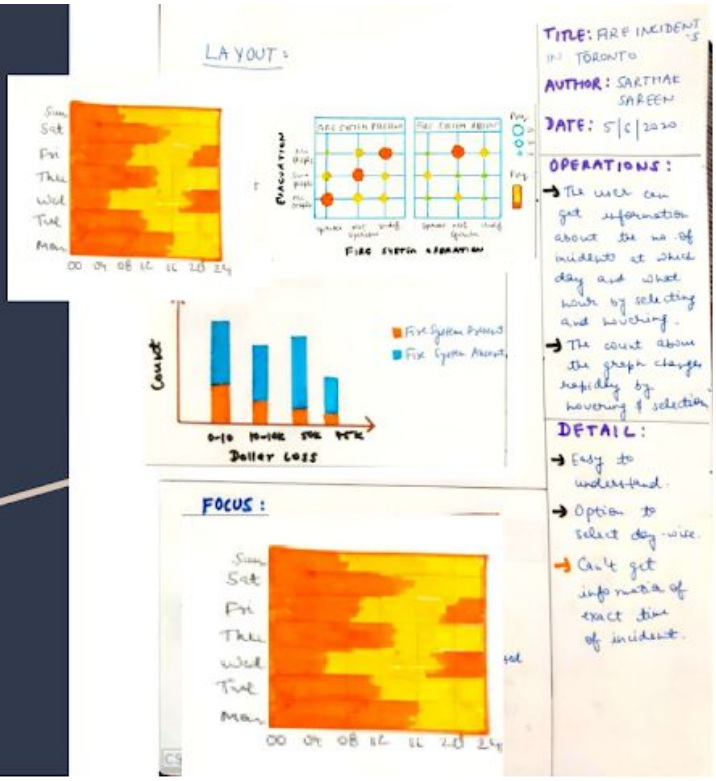
BRAINSTORM



FIVE DESIGN SHEET

SHEET 2

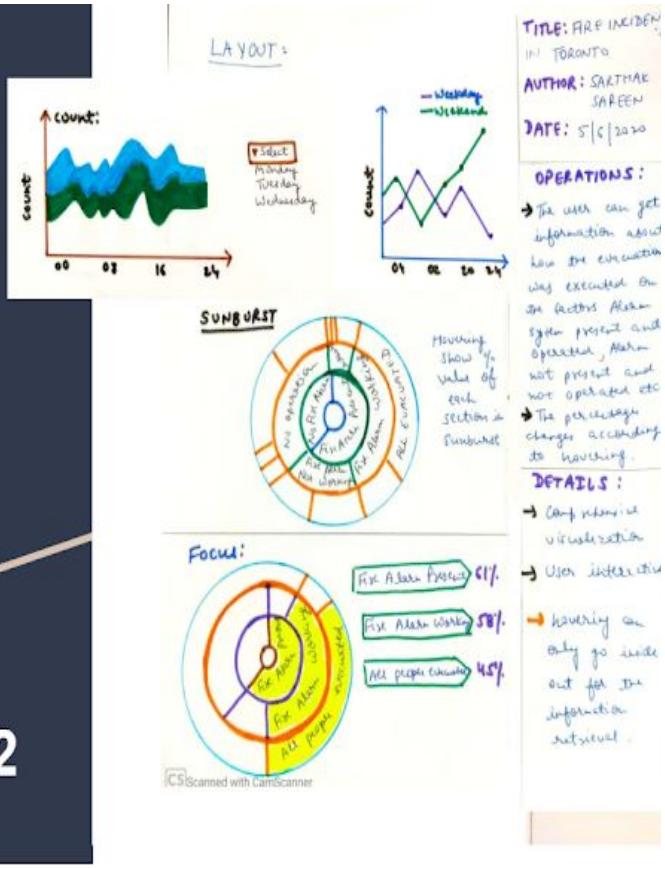
INITIAL DESIGN 1



FIVE DESIGN SHEET

SHEET 3

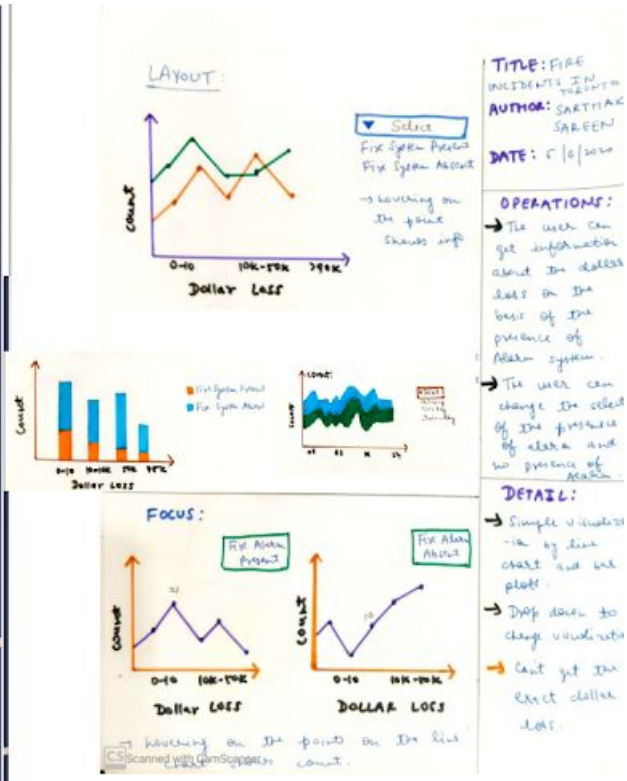
INITIAL DESIGN 2



FIVE DESIGN SHEET

SHEET 4

INITIAL DESIGN 3



FIVE DESIGN SHEET

SHEET 5

REALISATION