THE TRAGEDY OF FLIGHT: A COMPREHENSIVE CRASH ANALYSIS

TEAM MEMBERS:

M. VASU(HEAD OF THE TEAM)

P. KOKILAVANI

D. VARSHA

B. BARKAVI

INTRODUCTION:

* Air accident is caused due to pilot error, mechanical error, bed weather, sabotages or human

error. These error can happened anytime. Air accidents is also called flying mischance cases have a Great degree complexity. Aircraft crashes cause harm to the population as it may lead to the deaths of people or may even cause injuries to the people3. Crash investigations is a major research area and the major techniques used for this investigation are statistics. Grid computing, cloud computing, digital image processing and data mining with data mining, we can parse through extremely large amount of data and find out unknown pattern of air mischance. The Major of our project is to use data mining techniques to find out unknown patterns in the international flight crash datasets, the research is carried in aircraft crash and fatalities data collected from the year 1908 to 2009. This work is carried out using k-mean clustering data mining technique and cosine similarity measure.

The paper is further structured as follows theoretical background, methodology, experimental setup, analysis and conclusion Aircraft crashes are fundamental to weather (thunderstorm, strong winds, etc.), pilot error, etc. There is a necessity to study these issues. The study helps to understand who all are suffering, what is the reason for the crash and what harm is it causing to the general population

and how can we prevent them. The crashes may cause loss of
human life as well as loss of natural resources.

THEORETICAL BACKGROUND:

*The most common reasons for plane accidents are pilot
error, mechanical failure, human error, etc. These parameters are explained as follows:

1. Pilot Error – Roughly 50% of the aircraft losses incur due to pilot error. There are many chances for The pilots to cause errors from failing to program correctly to miscalculation of the required fuel.

2. Mechanical Failure -Despite developments in model and manufacturing standards of the aircraft, mechanical failures account for 20% of aircraft losses.

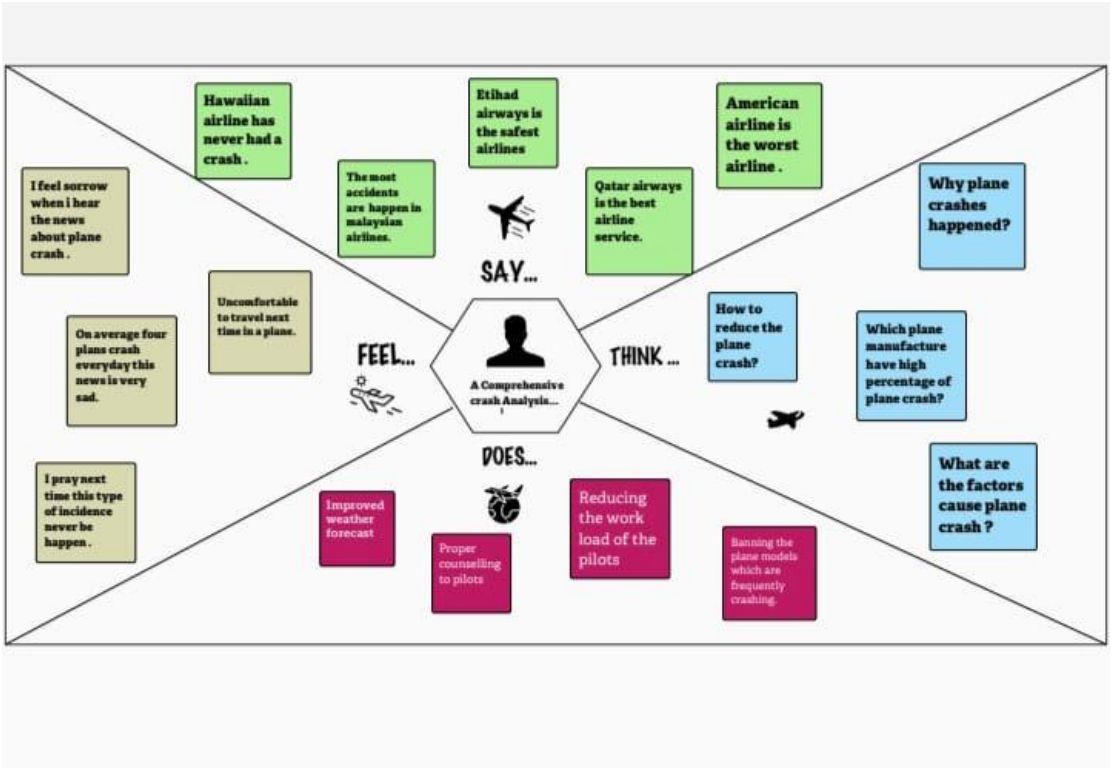
3. Weather - Despite of having multiple electronic aids, aircraft still struggle to function properly when the weather turns out to be unpleasant like in storms, snow and fog.

4. Sabotage - The dangers posed by sabotage are much less than many people seem to believe. Approximately 10% of aircraft losses occur due to sabotage.

5. Human Error – Mistakes can be made by humans operating when required to work for longer hours. Air traffic controllers, dispatchers, loaders, etc are some of the jobs that are operated by humans.

PROBLEM DEFINITION & DESIGN THINKING:

EMPATHY MAP:



IDEATUON & BRAINSTROMING MAP:


Fly on nonstop rountings
lesson to the flight attendants
Govt attention avoid the subtage
Pre checking of weather condition
let the flight attendant pour your hot drinks
keep your seat belt fastened while you are seated
Pre checking of the grounding path
Choose larger aircraft
Pay attention to the preflight briefing
Keep the overhead storage bin free of heavy articles
Pre checking of plane control
Keep your wits about you
Do not drink too much
Importance
Feasibility

RESULT:

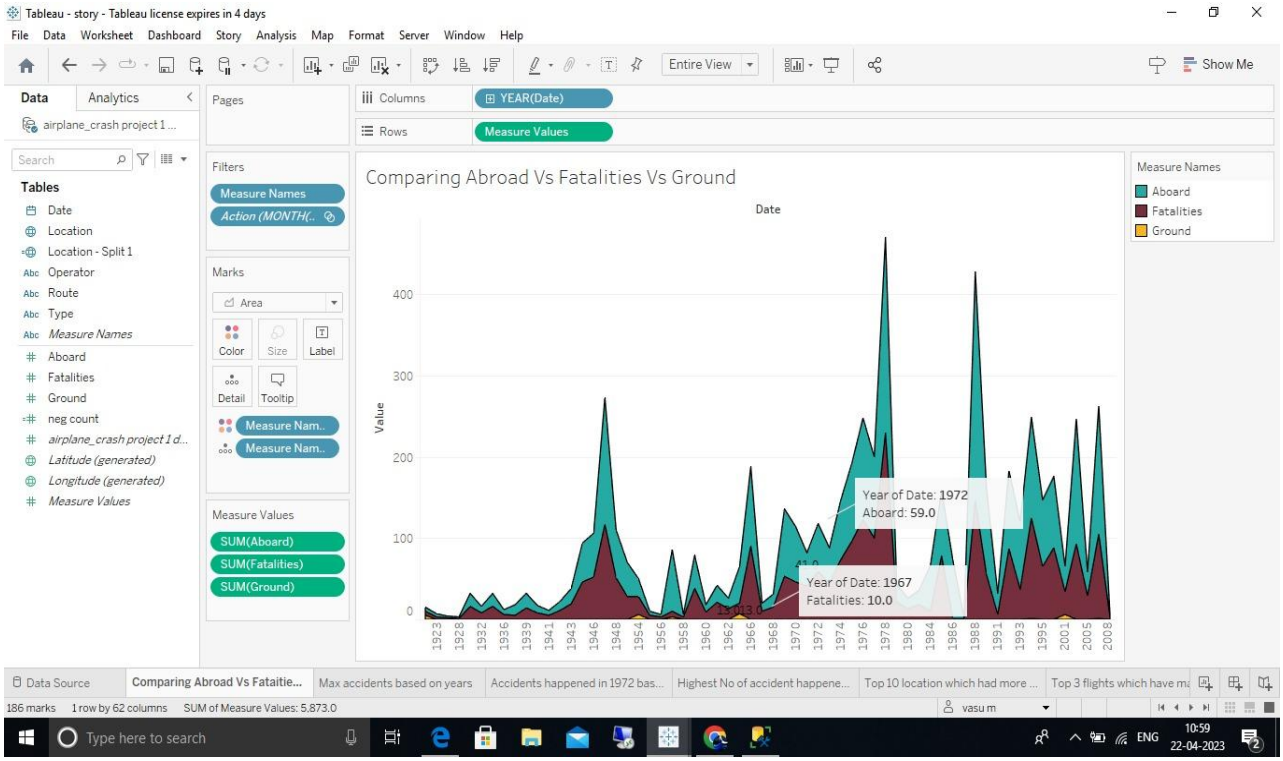


Tableau - story - Tableau license expires in 4 days
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Entire View
Show Me
Data
Analytics
airplane_crash project 1 ...
Search
Tables
Date
Location
Location - Split 1
Operator
Route
Type
Measure Names
Aboard
Fatalities
Ground
neg count
airplane_crash project 1 d...
Latitude (generated)
Longitude (generated)
Measure Values
Pages
Filters
Measure Names
Action (MONTH(..
Marks
Area
Color
Size
Label
Detail
Tooltip
Measure Nam..
Measure Nam..
Measure Values
SUM(Aboard)
SUM(Fatalities)
SUM(Ground)
Columns
YEAR(Date)
Rows
Measure Values
Comparing Abroad Vs Fatalities Vs Ground
Date
Value
Year of Date: 1972
Aboard: 59.0
Year of Date: 1967
Fatalities: 10.0
Measure Names
Aboard
Fatalities
Ground
Data Source
Comparing Abroad Vs Fataitie...
Max accidents based on years
Accidents happened in 1972 bas...
Highest No of accident happene...
Top 10 location which had more ...
Top 3 flights which have ma
186 marks
1 row by 62 columns
SUM of Measure Values: 5,873.0
vasu m
Type here to search
ENG
10:59
22-04-2023

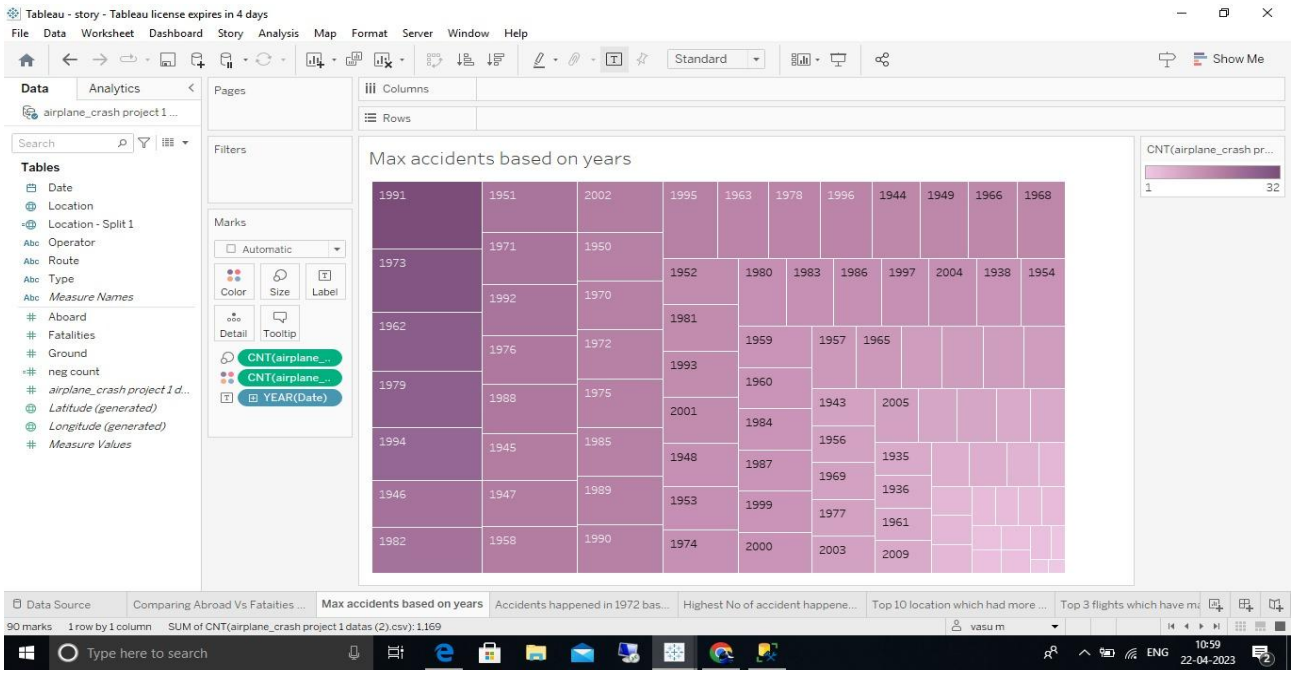


Tableau - story - Tableau license expires in 4 days
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Show Me
Data
Analytics
airplane_crash project 1 ...
Search
Tables
Date
Location
Location - Split 1
Operator
Route
Type
Measure Names
Aboard
Fatalities
Ground
neg count
airplane_crash project 1 d...
Latitude (generated)
Longitude (generated)
Measure Values
Pages
Filters
Marks
Automatic
Color
Size
Label
Detail
Tooltip
CNT(airplane_...
CNT(airplane_...
YEAR(Date)
Columns
Rows
Max accidents based on years
1991
1973
1962
1979
1994
1946
1982
1951
1971
1992
1976
1988
1945
1947
1958
2002
1950
1970
1972
1975
1985
1989
1990
1995
1963
1978
1996
1944
1949
1966
1968
1952
1980
1983
1986
1997
2004
1938
1954
1981
1959
1957
1965
1993
1960
1943
2005
2001
1984
1956
1935
1948
1987
1969
1936
1953
1999
1977
1961
1974
2000
2003
2009
CNT(airplane_pr...
1
32
Data Source
Comparing Abroad Vs Fataities ...
Max accidents based on years
Accidents happened in 1972 bas...
Highest No of accident happene...
Top 10 location which had more ...
Top 3 flights which have ma
90 marks
1 row by 1 column
SUM of CNT(airplane_crash project 1 datas (2).csv): 1,169
vasu m
Type here to search
ENG
10:59
22-04-2023

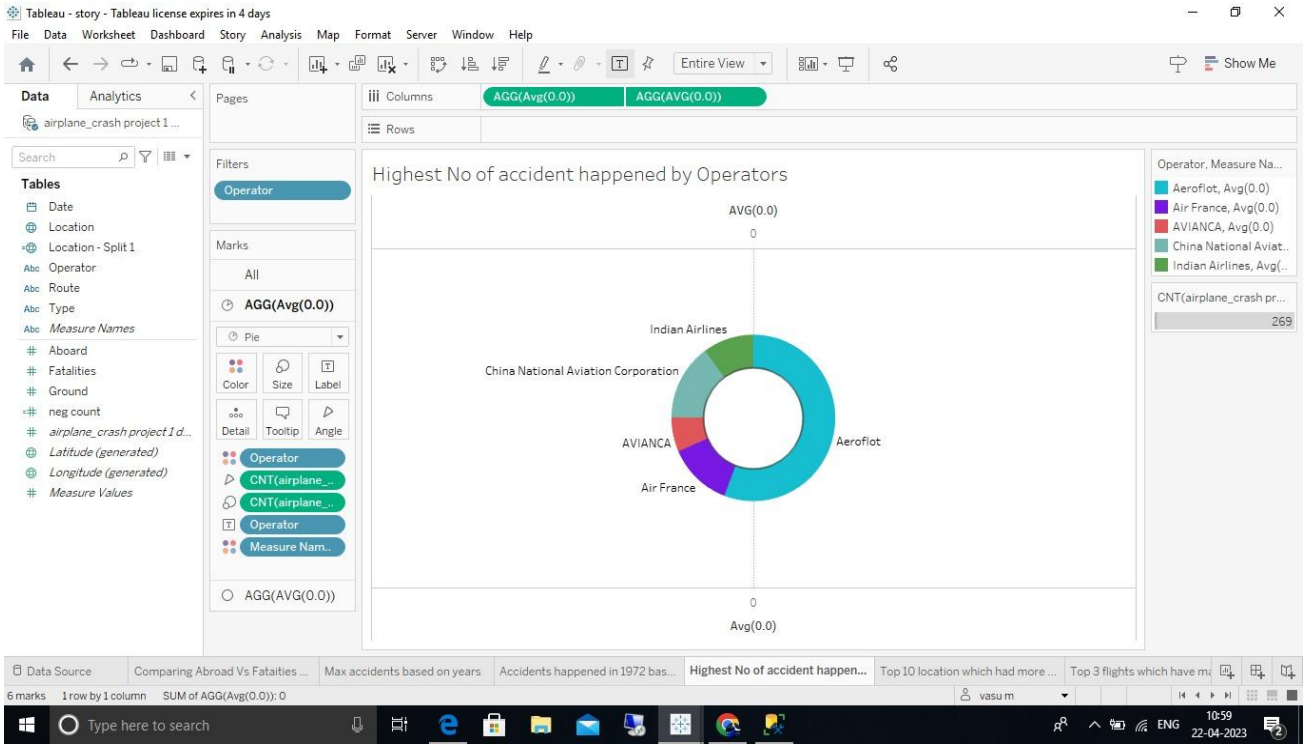


Tableau - story - Tableau license expires in 4 days
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Entire View
Show Me
Data
Analytics
airplane_crash project 1 ...
Search
Tables
Date
Location
Location - Split 1
Operator
Route
Type
Measure Names
Aboard
Fatalities
Ground
neg count
airplane_crash project 1 d...
Latitude (generated)
Longitude (generated)
Measure Values
Pages
Filters
Operator
Marks
All
AGG(Avg(0.0))
Pie
Color
Size
Label
Detail
Tooltip
Angle
Operator
CNT(airplane_...
CNT(airplane_...
Operator
Measure Nam..
AGG(AVG(0.0))
Columns
AGG(Avg(0.0))
AGG(AVG(0.0))
Rows
Highest No of accident happened by Operators
AVG(0.0)
0
Indian Airlines
China National Aviation Corporation
AVIANCA
Air France
Aeroflot
0
Avg(0.0)
Operator, Measure Na...
Aeroflot, Avg(0.0)
Air France, Avg(0.0)
AVIANCA, Avg(0.0)
China National Aviat..
Indian Airlines, Avg(..
CNT(airplane_crash pr...
269
Data Source
Comparing Abroad Vs Fatalities ...
Max accidents based on years
Accidents happened in 1972 bas...
Highest No of accident happen...
Top 10 location which had more ...
Top 3 flights which have ma
6 marks
1 row by 1 column
SUM of AGG(Avg(0.0)): 0
vasu m
Type here to search
ENG
10:59
22-04-2023

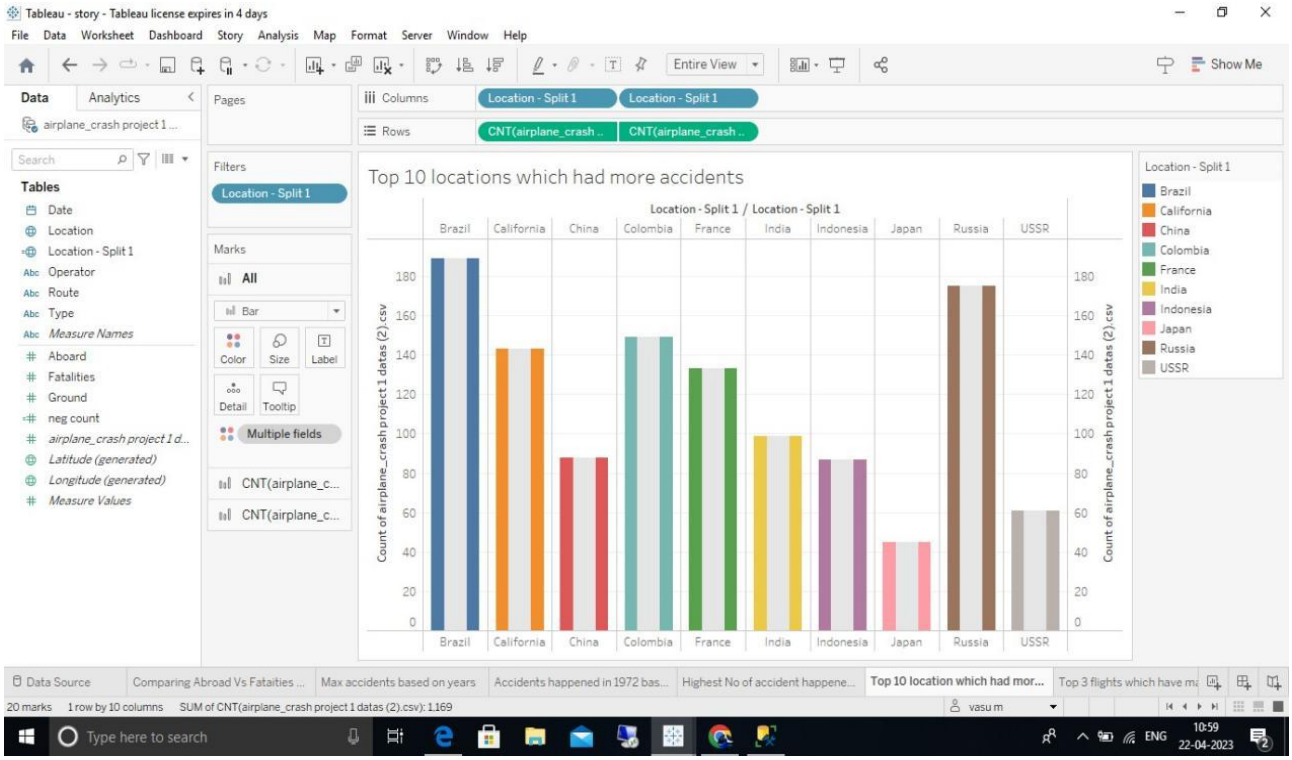


Tableau - story - Tableau license expires in 4 days
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Entire View
Show Me
Data
Analytics
airplane_crash project 1 ...
Search
Tables
Date
Location
Location - Split 1
Operator
Route
Type
Measure Names
Aboard
Fatalities
Ground
neg count
airplane_crash project 1 d...
Latitude (generated)
Longitude (generated)
Measure Values
Pages
Filters
Location - Split 1
Marks
All
Bar
Color
Size
Label
Detail
Tooltip
Multiple fields
CNT(airplane_c...
CNT(airplane_c...
Columns
Location - Split 1
Location - Split 1
Rows
CNT(airplane_crash ..
CNT(airplane_crash ..
Top 10 locations which had more accidents
Location - Split 1 / Location - Split 1
Brazil
California
China
Colombia
France
India
Indonesia
Japan
Russia
USSR
Count of airplane_crash project 1 datas (2).csv
0
20
40
60
80
100
120
140
160
180
Location - Split 1
Brazil
California
China
Colombia
France
India
Indonesia
Japan
Russia
USSR
Data Source
Comparing Abroad Vs Fatalities ...
Max accidents based on years
Accidents happened in 1972 bas...
Highest No of accident happene...
Top 10 location which had mor...
Top 3 flights which have ma
20 marks
1 row by 10 columns
SUM of CNT(airplane_crash project 1 datas (2).csv): 1169
vasu m
Type here to search
ENG
10:59
22-04-2023

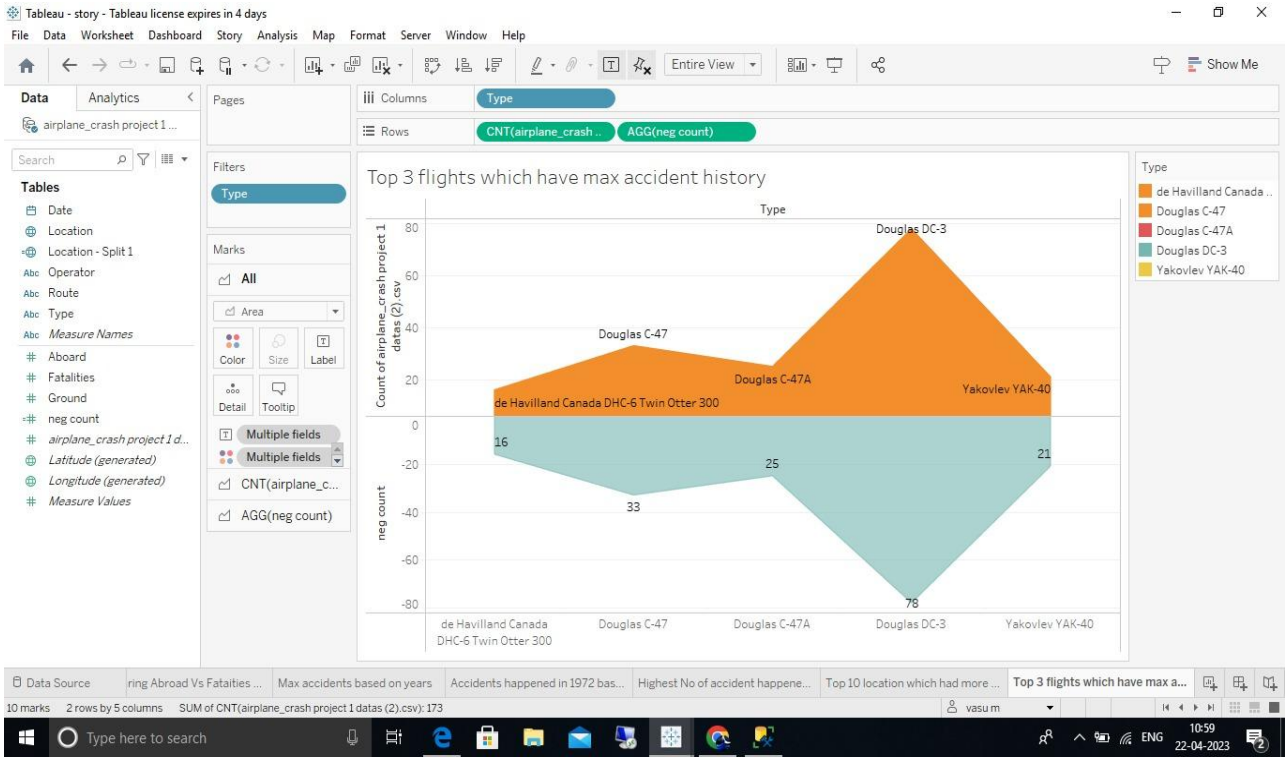


Tableau - story - Tableau license expires in 4 days
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Entire View
Show Me
Data
Analytics
airplane_crash project 1 ...
Tables
Date
Location
Location - Split 1
Operator
Route
Type
Measure Names
Aboard
Fatalities
Ground
neg count
airplane_crash project 1 d...
Latitude (generated)
Longitude (generated)
Measure Values
Pages
Filters
Type
Marks
All
Area
Color
Size
Label
Detail
Tooltip
Multiple fields
Multiple fields
CNT(airplane_c...
AGG(neg count)
Columns
Type
Rows
CNT(airplane_crash ..
AGG(neg count)
Top 3 flights which have max accident history
Type
Count of airplane_crash project 1 datas (2).csv
neg count
Douglas DC-3
Douglas C-47
Douglas C-47A
Yakovlev YAK-40
de Havilland Canada DHC-6 Twin Otter 300
16
33
25
78
21
de Havilland Canada DHC-6 Twin Otter 300
Douglas C-47
Douglas C-47A
Douglas DC-3
Yakovlev YAK-40
Type
de Havilland Canada ..
Douglas C-47
Douglas C-47A
Douglas DC-3
Yakovlev YAK-40
Data Source
ring Abroad Vs Fatalities ...
Max accidents based on years
Accidents happened in 1972 bas...
Highest No of accident happene...
Top 10 location which had more ...
Top 3 flights which have max a...
10 marks
2 rows by 5 columns
SUM of CNT(airplane_crash project 1 datas (2).csv): 173
vasu m
Type here to search
ENG
10:59
22-04-2023

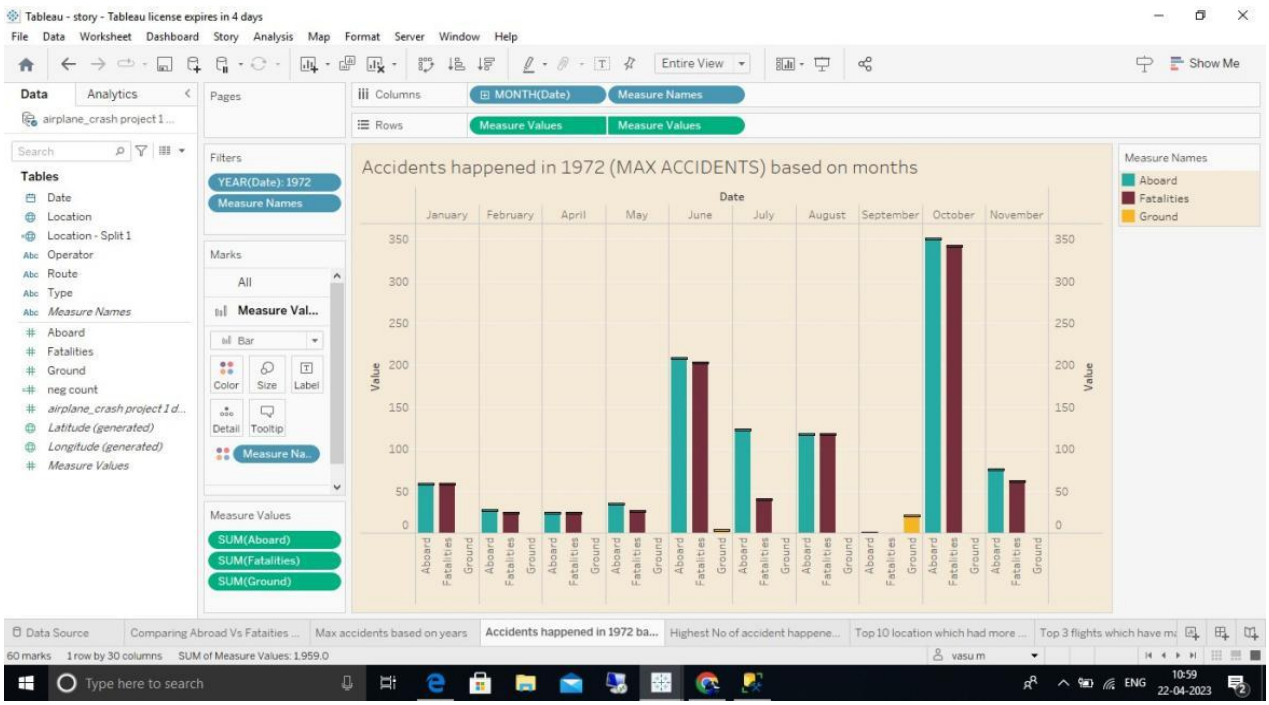
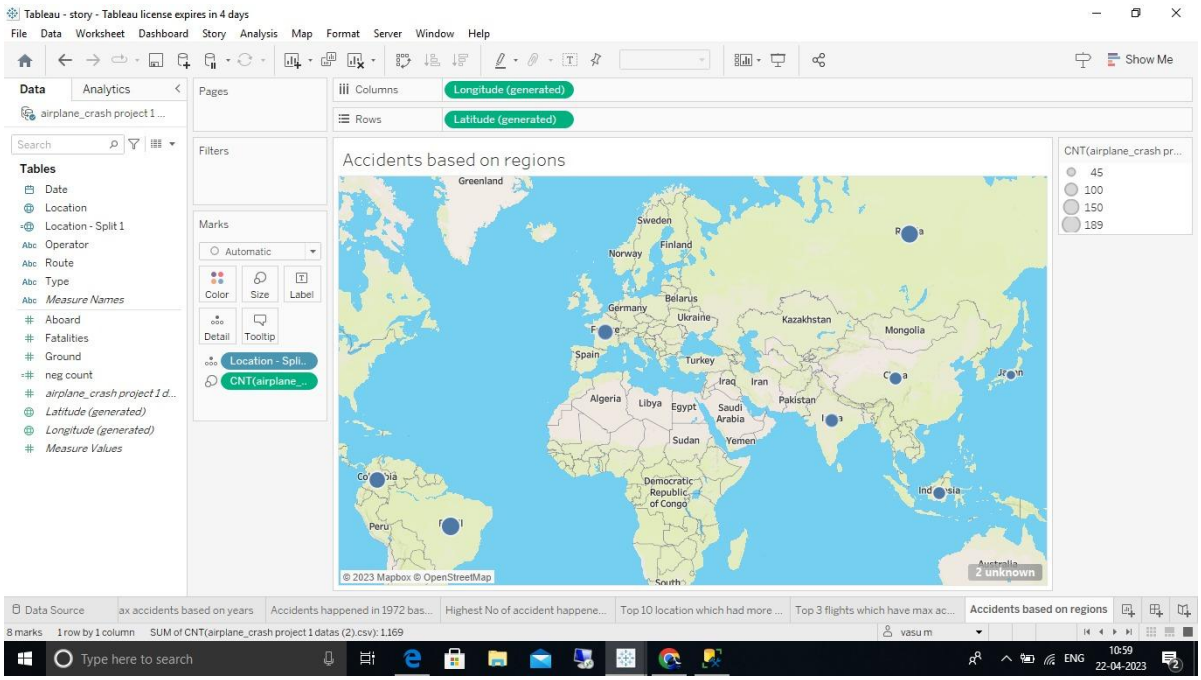


Tableau - story - Tableau license expires in 4 days
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Entire View
Show Me
Data
Analytics
airplane_crash project 1 ...
Tables
Date
Location
Location - Split 1
Operator
Route
Type
Measure Names
Aboard
Fatalities
Ground
neg count
airplane_crash project 1 d...
Latitude (generated)
Longitude (generated)
Measure Values
Pages
Filters
YEAR(Date): 1972
Measure Names
Marks
All
Measure Val...
Bar
Color
Size
Label
Detail
Tooltip
Measure Na...
Measure Values
SUM(Aboard)
SUM(Fatalities)
SUM(Ground)
Columns
MONTH(Date)
Measure Names
Rows
Measure Values
Measure Values
Accidents happened in 1972 (MAX ACCIDENTS) based on months
Date
January
February
April
May
June
July
August
September
October
November
Value
Measure Names
Aboard
Fatalities
Ground
Data Source
Comparing Abroad Vs Fatalities ...
Max accidents based on years
Accidents happened in 1972 ba...
Highest No of accident happene...
Top 10 location which had more ...
Top 3 flights which have ma
60 marks
1 row by 30 columns
SUM of Measure Values: 1,959.0
vasu m
Type here to search
ENG
10:59
22-04-2023



LITERATURE SURVEY :

*Flight Crash Investigation Using Data Mining Techniques". The research work is done for identifying aboard/ground fatality rate with operators and locations well as to find similarity among the plane crashes [3].

*Analyzing relationship and incidents. In proceedings of the international conference on research in air transportation. In this research he employed a data mining technique to conduct the holistic analysis of aircraft incident data in relation to the accident data. The analysis

identifies the relation between the accident and incident data and finds the patterns of casual and contributory factors which are significantly evocatively with the aircraft accident [4].

*Improvement of aircraft accident investigation through expert systems. In this research presented in this paper shown that expert system methodology is a robust approach to analyzing the aircraft accident investigation [3].

*Data clustering algorithm applications CRC press. In this research the data clustering algorithm and application provides complete coverage of area of clustering from the basic method to more refined and data clustering approaches.it pays special attention to the in graphs, social networks [6].

METHODOLOGY :

* In allocation natural language processing latent Dirichlet (LDA)is a generative statistical model that allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar.

CONCLUSION:

*The K-Mean clustering technique was used to find the clusters and fatality for the flight crash investigation. The fatality of ground is more than aboard. Aeroflot has the maximum fatality,I.e. 4266 all the other operators. The route Terrine - Las Pal-mas has got the maximum number of fatalities. The research work can be extended using other clustering techniques like Density Based, Hierarchical clustering. The summary report of the datasets is used to identify better clusters using distance measures like cosine similarity. Cosine similarity is used for finding the similarity among the crashes. Majority is used for finding crashed in USA.

The flights with id This study covered accidents and incidents pertaining to commercial flights within the United States. A similar study could be conducted on the General Aviation (GA). Depending on the availability of the data, the studies could be extended to regions in other countries.

FUTURE SCOPE:

*Transmission - Top mounted antenna communicate with satellites at higher altitudes, at lower altitudes, data can be sent directly to the ground receivers.

*Bandwidth-To save money, aircraft could flash data intermittently, switching to streaming ways in an emergency.

*Satellite - By 2015, a castellations of desiccate search towards technology, rescue satellites will track aeroplane exact location more quickly globally.

*Storage Servers-
Today uncountable servers
around the world stores & send real time flights recorded data. Eventually, in coming future every airline would have its own server.

APPENDIX:

Source code:

https://drive.google.com/file/d/1--5goHu7iDrtr_M8u0eTIMFHcUultQm/view?usp=drivesdk