Introduction to Data Management PROJECT REPORT

(Project Semester August-December 2021)

PROJECT REPORT ON

Aviation Accident Database

Submitted by

Shahrukh Zeya

12002492

Program: Bachelor of Technology Section: KM008
Course Code: INT217

Under the Guidance of
Komal Arora:
17783
Assistant Professor

Discipline of CSE/IT

Lovely School of Computer Science & Engineering
Lovely Professional University, Phagwara

DECLARATION

I **Shahrukh Zeya**, student of Computer Science & Engineering under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declarethat all the information furnished in this project report is based on my own intensive-work and is genuine.

Date: December 16, 2021

Shahrukh Zeya

Registration No:
12002492

ACKNOWLEDGEMENT

A project work is a combination of views, ideas, suggestions and contribution of many people. Thus, one of the pleasant parts of writing the report is to thank those who have contributed towards its fulfilment.

LI consider it as great privilege to have esteemed Lecturer **Ms. Komal Arora** as my project guide. I take this opportunity to express my sincere gratitude to him through constant advice and constructive criticism nourished my interest in the subject and provided a free and pleasant atmosphere to work against all odd situations. I avail this opportunity to extend my heart full thanks and deep respect to faculty member for their able guidance during this project.

My gratitude to all those, who responded to my questionnaire in a well-defined manner and helped me acquiring knowledge.

I would like to communicate a deep sense of gratitude to all these people without whom my project would not have been completed, such a great learning experience.

Shahrukh Zeya

CONTENTS

INTRODUCTION

Data management is the practice of collecting, keeping, and using data securely, efficiently, and cost-effectively. The goal of data management is to help people, organizations, and connected things optimize the use of data within the bounds of policy and regulation so that they can make decisions and take actions that maximize the benefit to the organization. A robust data management strategy is becoming more important than ever as organizations increasingly rely on intangible assets to create value.

Data Analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision- making.

This project is based on analysis on Accident Aviation database.

The NTSB aviation accident database contains information from 1962 and later about civil aviation accidents and selected incidents within the United States, its territories and possessions, and in international waters.

OBJECTIVES/SCOPE OF ANALYSIS

After analysis of the dataset, I found these five Objective for this project to be solved.

- Country wise total accident segregated by total minor accident and total serious injuries.
- Aviation condition after accidents.
- Total accidents according to types of engine.
- Country-wise total number of accidents of different purpose of flights.
- Total number of accidents segregated by year and month.

SOURCE OF DATASET:

Source data set link:
https://www.kaggle.com/khsamaha/aviation-accident-database-synopses

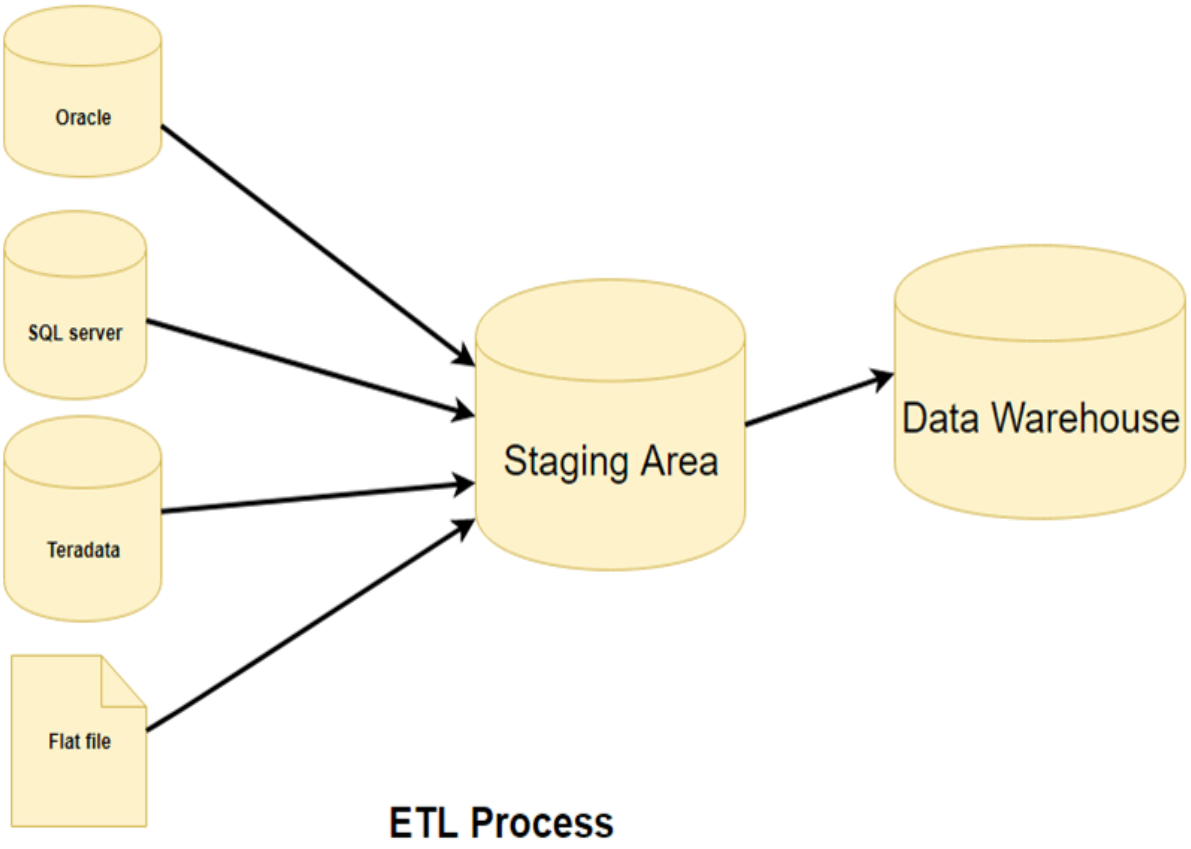
These are the following field that are available in the Aviation accident dataset.

- Event.Id
- Investigation.Type
- Accident.Number
- Event.Date
- LOCATION
- Country
- Latitude
- Longitude
- Airport.Code
- Airport.Name
- Injury.Severity
- Aircraft.damage
- Aircraft.Category
- Registration.Number
- Make
- Model
- Amateur.Built
- Number.of.Engines
- Engine.Type
- FAR.Description
- Schedule
- Purpose.of.flight
- Air.carrier

- Total.Fatal.Injuries
- Total.Serious.Injuries
- Total.Minor.Injuries
- Total.Uninjured
- Weather.Condition

ETL Process:

ETL is a process in Data Warehousing and it stands for Extract, Transform and Load. It is a process in which an ETL tool extracts the data from various data source systems, transforms it in the staging area, and then finally, loads it into the Data Warehouse system.



ETL Process

Let us understand each step of the ETL process in-depth:

Extraction:
The first step of the ETL process is extraction. In this step, data from various source systems is extracted which can be in various formats like relational databases, No SQL, XML, and flat files into the staging area. It is important to extract the data from various source systems and store it into the staging area first and not directly into the data

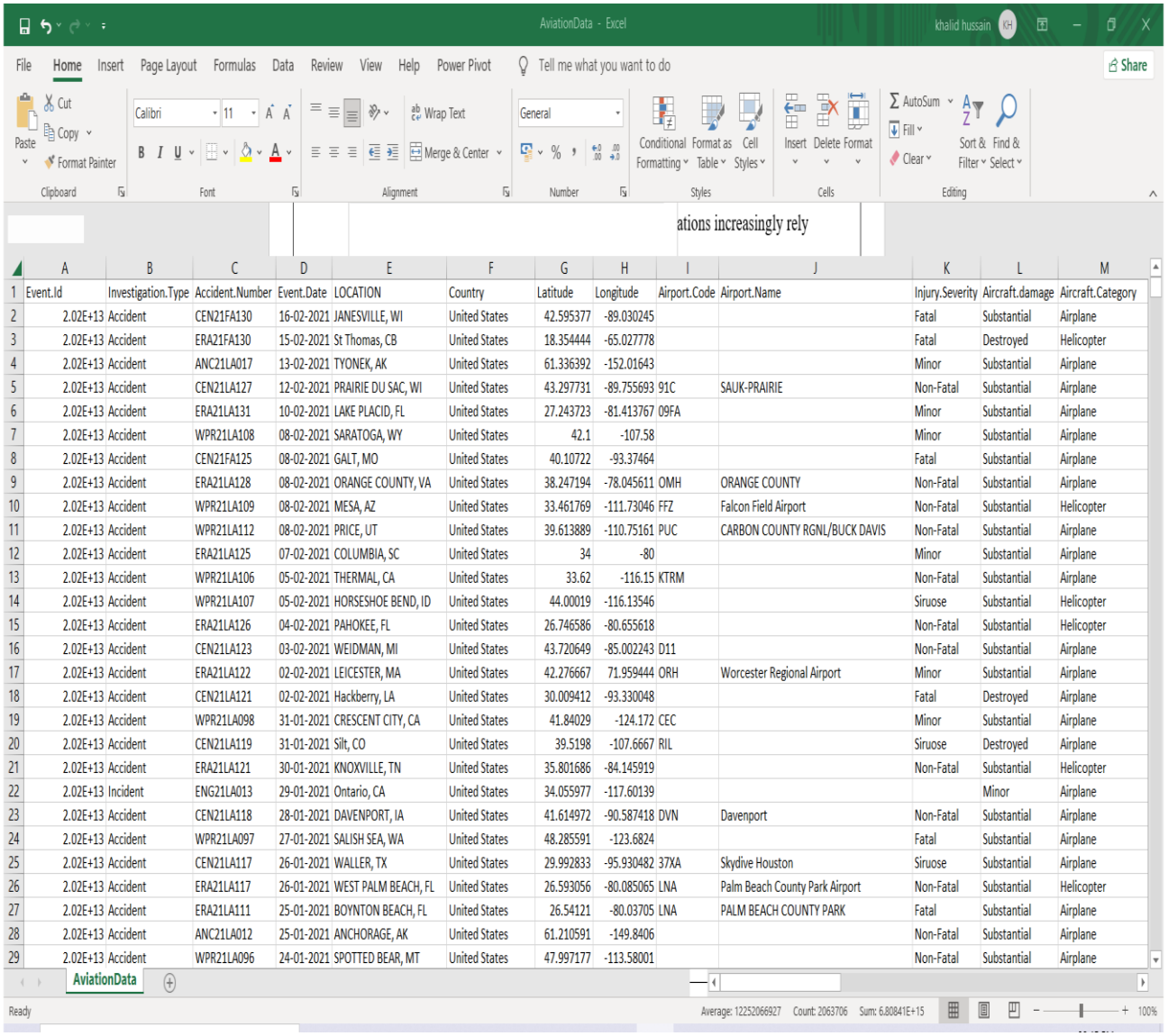
warehouse because the extracted data is in various formats and can be corrupted also. Hence loading it directly into the data warehouse may damage it and rollback will be much more difficult. Therefore, this is one of the most important steps of ETL process.

Transformation:
The second step of the ETL process is transformation. In this step, a set of rules or functions are applied on the extracted data to convert it into a single standard format. It may involve following processes/tasks:

- **Filtering** – loading only certain attributes into the data warehouse.
- **Cleaning –** filling up the NULL values with some default values, mapping U.S.A, United States, and America into USA, etc.
- **Joining –** joining multiple attributes into one.
- **Splitting –** splitting a single attribute into multiple attributes.
- **Sorting –** sorting tuples on the basis of some attribute (generally key-attribute).

Loading:
The third and final step of the ETL process is loading. In this step, the transformed data is finally loaded into the data warehouse. Sometimes the data is updated by loading into the data warehouse very frequently and sometimes it is done after longer but regular intervals. The rate and period of loading solely depends on the requirements and varies from system to system.

Initially the raw dataset was look like as shown in given picture:



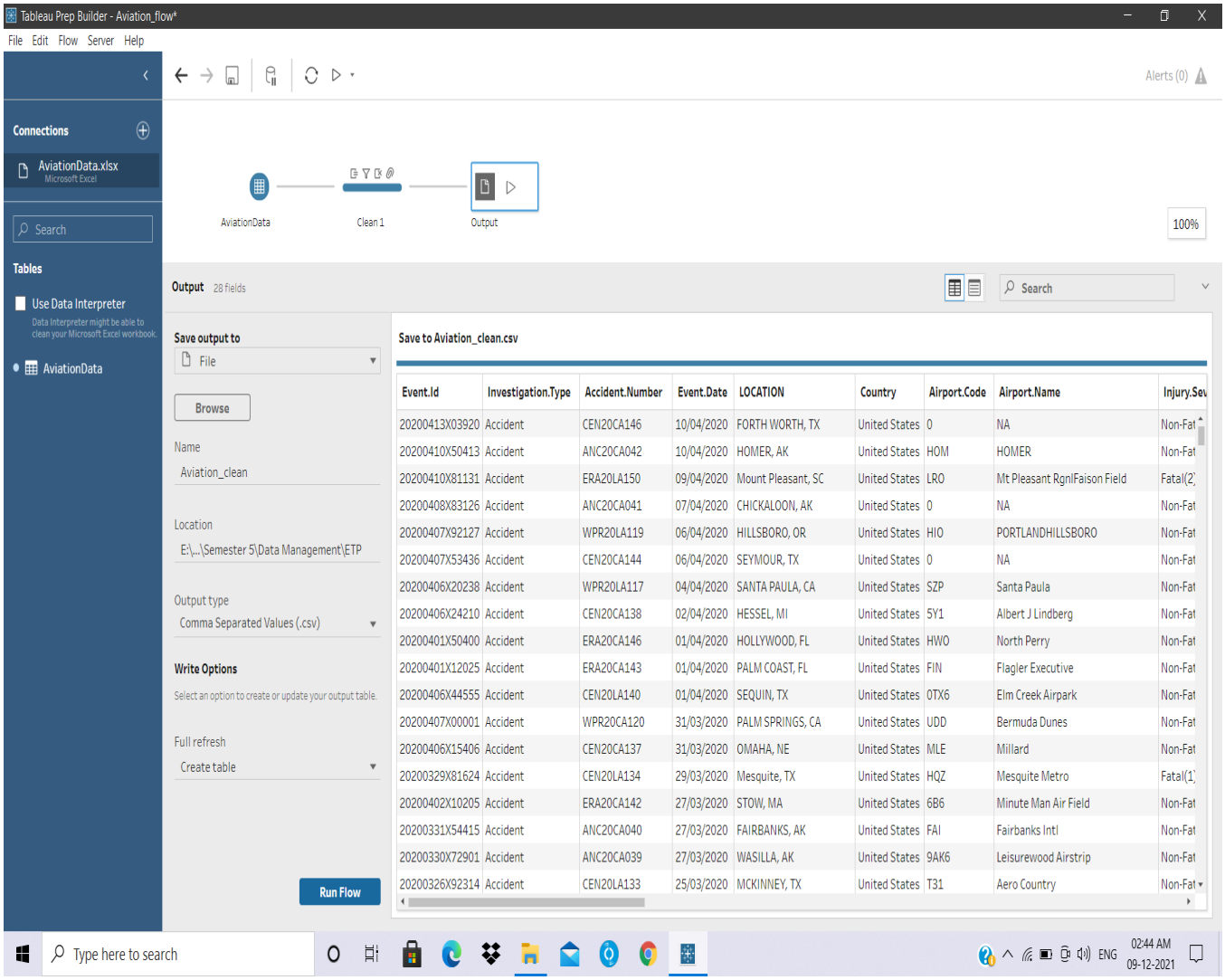
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Event.Id	Investigation.Type	Accident.Number	Event.Date	LOCATION	Country	Latitude	Longitude	Airport.Code	Airport.Name	Injury.Severity	Aircraft.damage	Aircraft.Category
2	2.02E+13	Accident	CEN21FA130	16-02-2021	JANESVILLE, WI	United States	42.595377	-89.030245			Fatal	Substantial	Airplane
3	2.02E+13	Accident	ERA21FA130	15-02-2021	St Thomas, CB	United States	18.354444	-65.027778			Fatal	Destroyed	Helicopter
4	2.02E+13	Accident	ANC21LA017	13-02-2021	TYONEK, AK	United States	61.336392	-152.01643			Minor	Substantial	Airplane
5	2.02E+13	Accident	CEN21LA127	12-02-2021	PRAIRIE DU SAC, WI	United States	43.297731	-89.755693	91C	SAUK-PRAIRIE	Non-Fatal	Substantial	Airplane
6	2.02E+13	Accident	ERA21LA131	10-02-2021	LAKE PLACID, FL	United States	27.243723	-81.413767	09FA		Minor	Substantial	Airplane
7	2.02E+13	Accident	WPR21LA108	08-02-2021	SARATOGA, WY	United States	42.1	-107.58			Minor	Substantial	Airplane
8	2.02E+13	Accident	CEN21FA125	08-02-2021	GALT, MO	United States	40.10722	-93.37464			Fatal	Substantial	Airplane
9	2.02E+13	Accident	ERA21LA128	08-02-2021	ORANGE COUNTY, VA	United States	38.247194	-78.045611	OMH	ORANGE COUNTY	Non-Fatal	Substantial	Airplane
10	2.02E+13	Accident	WPR21LA109	08-02-2021	MESA, AZ	United States	33.461769	-111.73046	FFZ	Falcon Field Airport	Non-Fatal	Substantial	Helicopter
11	2.02E+13	Accident	WPR21LA112	08-02-2021	PRICE, UT	United States	39.613889	-110.75161	PUC	CARBON COUNTY RGNL/BUCK DAVIS	Non-Fatal	Substantial	Airplane
12	2.02E+13	Accident	ERA21LA125	07-02-2021	COLUMBIA, SC	United States	34	-80			Minor	Substantial	Airplane
13	2.02E+13	Accident	WPR21LA106	05-02-2021	THERMAL, CA	United States	33.62	-116.15	KTRM		Non-Fatal	Substantial	Airplane
14	2.02E+13	Accident	WPR21LA107	05-02-2021	HORSESHOE BEND, ID	United States	44.00019	-116.13546			Siruose	Substantial	Helicopter
15	2.02E+13	Accident	ERA21LA126	04-02-2021	PAHOKEE, FL	United States	26.746586	-80.655618			Non-Fatal	Substantial	Helicopter
16	2.02E+13	Accident	CEN21LA123	03-02-2021	WEIDMAN, MI	United States	43.720649	-85.002243	D11		Non-Fatal	Substantial	Airplane
17	2.02E+13	Accident	ERA21LA122	02-02-2021	LEICESTER, MA	United States	42.276667	71.959444	ORH	Worcester Regional Airport	Minor	Substantial	Airplane
18	2.02E+13	Accident	CEN21LA121	02-02-2021	Hackberry, LA	United States	30.009412	-93.330048			Fatal	Destroyed	Airplane
19	2.02E+13	Accident	WPR21LA098	31-01-2021	CRESCENT CITY, CA	United States	41.84029	-124.172	CEC		Minor	Substantial	Airplane
20	2.02E+13	Accident	CEN21LA119	31-01-2021	Silt, CO	United States	39.5198	-107.6667	RIL		Siruose	Destroyed	Airplane
21	2.02E+13	Accident	ERA21LA121	30-01-2021	KNOXVILLE, TN	United States	35.801686	-84.145919			Non-Fatal	Substantial	Helicopter
22	2.02E+13	Incident	ENG21LA013	29-01-2021	Ontario, CA	United States	34.055977	-117.60139				Minor	Airplane
23	2.02E+13	Accident	CEN21LA118	28-01-2021	DAVENPORT, IA	United States	41.614972	-90.587418	DVN	Davenport	Non-Fatal	Substantial	Airplane
24	2.02E+13	Accident	WPR21LA097	27-01-2021	SALISH SEA, WA	United States	48.285591	-123.6824			Fatal	Substantial	Airplane
25	2.02E+13	Accident	CEN21LA117	26-01-2021	WALLER, TX	United States	29.992833	-95.930482	37XA	Skydive Houston	Siruose	Substantial	Airplane
26	2.02E+13	Accident	ERA21LA117	26-01-2021	WEST PALM BEACH, FL	United States	26.593056	-80.085065	LNA	Palm Beach County Park Airport	Non-Fatal	Substantial	Helicopter
27	2.02E+13	Accident	ERA21LA111	25-01-2021	BOYNTON BEACH, FL	United States	26.54121	-80.03705	LNA	PALM BEACH COUNTY PARK	Fatal	Substantial	Airplane
28	2.02E+13	Accident	ANC21LA012	25-01-2021	ANCHORAGE, AK	United States	61.210591	-149.8406			Non-Fatal	Substantial	Airplane
29	2.02E+13	Accident	WPR21LA096	24-01-2021	SPOTTED BEAR, MT	United States	47.997177	-113.58001			Non-Fatal	Substantial	Airplane

Preparing data using tabluea prep:

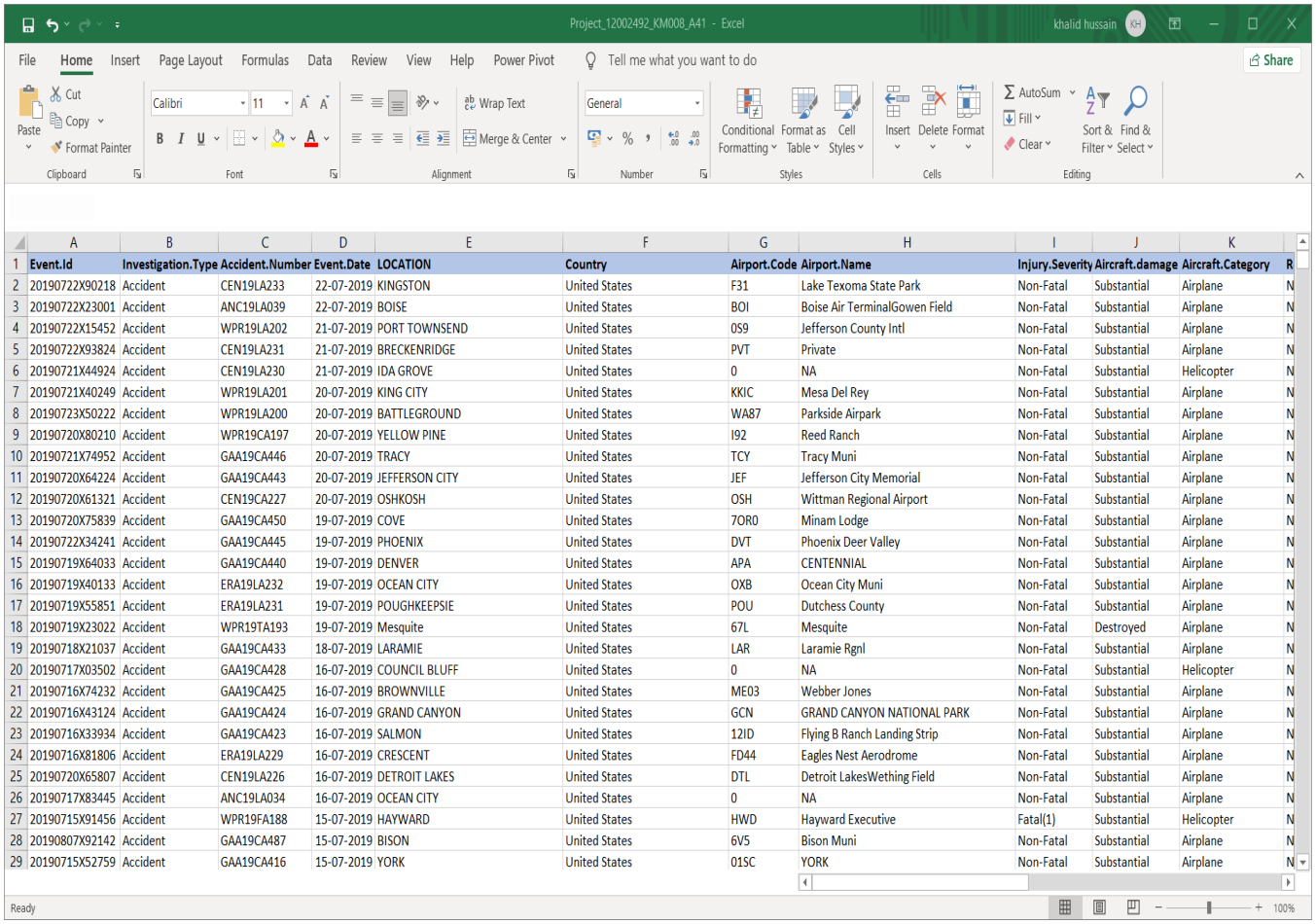
These are the following steps used in cleaning and preparing data in tabluea

- Removing Null value
- Changing data type
- Renaming field name
- Removing field with more than 50 percent null value
- Grouping common name
- Grouping Punctuation mistake

After applying all the above operations in the tableau prep finally, I run the output and save the file in CSV format.



After applying all the cleaning operation in the tableau prep, I imported the data in the CSV file format and after that apply some other operation in excel and finally now data base is ready to work on it, let's have a look how it looks like.



	A	B	C	D	E	F	G	H	I	J	K	
1	Event.Id	Investigation.Type	Accident.Number	Event.Date	LOCATION	Country	Airport.Code	Airport.Name	Injury.Severity	Aircraft.damage	Aircraft.Category	R
2	20190722X90218	Accident	CEN19LA233	22-07-2019	KINGSTON	United States	F31	Lake Texoma State Park	Non-Fatal	Substantial	Airplane	N
3	20190722X23001	Accident	ANC19LA039	22-07-2019	BOISE	United States	BOI	Boise Air TerminalGowen Field	Non-Fatal	Substantial	Airplane	N
4	20190722X15452	Accident	WPR19LA202	21-07-2019	PORT TOWNSEND	United States	0S9	Jefferson County Intl	Non-Fatal	Substantial	Airplane	N
5	20190722X93824	Accident	CEN19LA231	21-07-2019	BRECKENRIDGE	United States	PVT	Private	Non-Fatal	Substantial	Airplane	N
6	20190721X44924	Accident	CEN19LA230	21-07-2019	IDA GROVE	United States	0	NA	Non-Fatal	Substantial	Helicopter	N
7	20190721X40249	Accident	WPR19LA201	20-07-2019	KING CITY	United States	KKIC	Mesa Del Rey	Non-Fatal	Substantial	Airplane	N
8	20190723X50222	Accident	WPR19LA200	20-07-2019	BATTLEGROUND	United States	WA87	Parkside Airpark	Non-Fatal	Substantial	Airplane	N
9	20190720X80210	Accident	WPR19CA197	20-07-2019	YELLOW PINE	United States	I92	Reed Ranch	Non-Fatal	Substantial	Airplane	N
10	20190721X74952	Accident	GAA19CA446	20-07-2019	TRACY	United States	TCY	Tracy Muni	Non-Fatal	Substantial	Airplane	N
11	20190720X64224	Accident	GAA19CA443	20-07-2019	JEFFERSON CITY	United States	JEF	Jefferson City Memorial	Non-Fatal	Substantial	Airplane	N
12	20190720X61321	Accident	CEN19CA227	20-07-2019	OSHKOSH	United States	OSH	Wittman Regional Airport	Non-Fatal	Substantial	Airplane	N
13	20190720X75839	Accident	GAA19CA450	19-07-2019	COVE	United States	7OR0	Minam Lodge	Non-Fatal	Substantial	Airplane	N
14	20190722X34241	Accident	GAA19CA445	19-07-2019	PHOENIX	United States	DVT	Phoenix Deer Valley	Non-Fatal	Substantial	Airplane	N
15	20190719X64033	Accident	GAA19CA440	19-07-2019	DENVER	United States	APA	CENTENNIAL	Non-Fatal	Substantial	Airplane	N
16	20190719X40133	Accident	ERA19LA232	19-07-2019	OCEAN CITY	United States	OXB	Ocean City Muni	Non-Fatal	Substantial	Airplane	N
17	20190719X55851	Accident	ERA19LA231	19-07-2019	POUGHKEEPSIE	United States	POU	Dutchess County	Non-Fatal	Substantial	Airplane	N
18	20190719X23022	Accident	WPR19TA193	19-07-2019	Mesquite	United States	67L	Mesquite	Non-Fatal	Destroyed	Airplane	N
19	20190718X21037	Accident	GAA19CA433	18-07-2019	LARAMIE	United States	LAR	Laramie Rgnl	Non-Fatal	Substantial	Airplane	N
20	20190717X03502	Accident	GAA19CA428	16-07-2019	COUNCIL BLUFF	United States	0	NA	Non-Fatal	Substantial	Helicopter	N
21	20190716X74232	Accident	GAA19CA425	16-07-2019	BROWNVILLE	United States	ME03	Webber Jones	Non-Fatal	Substantial	Airplane	N
22	20190716X43124	Accident	GAA19CA424	16-07-2019	GRAND CANYON	United States	GCN	GRAND CANYON NATIONAL PARK	Non-Fatal	Substantial	Airplane	N
23	20190716X33934	Accident	GAA19CA423	16-07-2019	SALMON	United States	12ID	Flying B Ranch Landing Strip	Non-Fatal	Substantial	Airplane	N
24	20190716X81806	Accident	ERA19LA229	16-07-2019	CRESCENT	United States	FD44	Eagles Nest Aerodrome	Non-Fatal	Substantial	Airplane	N
25	20190720X65807	Accident	CEN19LA226	16-07-2019	DETROIT LAKES	United States	DTL	Detroit LakesWething Field	Non-Fatal	Substantial	Airplane	N
26	20190717X83445	Accident	ANC19LA034	16-07-2019	OCEAN CITY	United States	0	NA	Non-Fatal	Substantial	Airplane	N
27	20190715X91456	Accident	WPR19FA188	15-07-2019	HAYWARD	United States	HWD	Hayward Executive	Fatal(1)	Substantial	Helicopter	N
28	20190807X92142	Accident	GAA19CA487	15-07-2019	BISON	United States	6V5	Bison Muni	Non-Fatal	Substantial	Airplane	N
29	20190715X52759	Accident	GAA19CA416	15-07-2019	YORK	United States	01SC	YORK	Non-Fatal	Substantial	Airplane	N

Analysis of Dataset

1. Country wise total accident segregated by total minor accident and total serious injuries

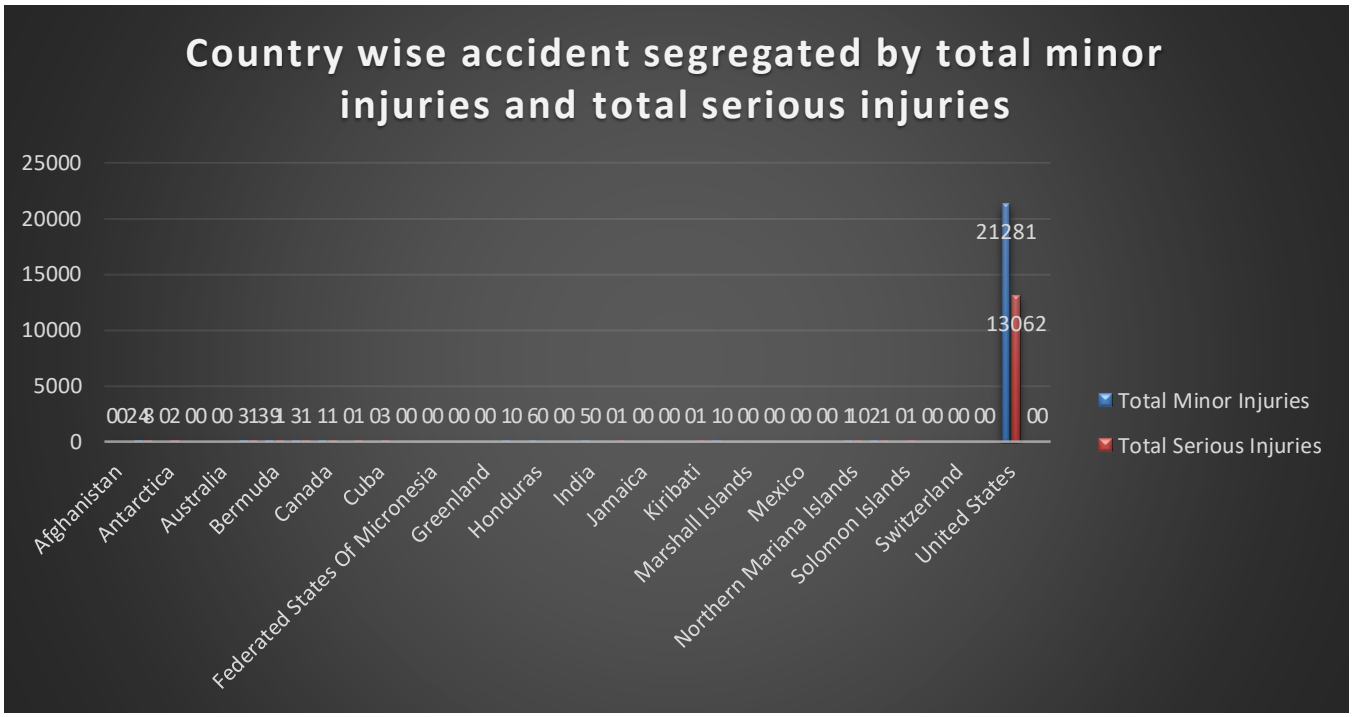
a) **Introduction:** The analysis shows that country wise total minor injury and total major injury.

b) Specific Requirements/Functions and Formulas:

i. Pivot table of country
ii. Pivot table of major injury
iii. Pivot table of minor injury
iv. Column Chart

c) Analysis Results:

• Analysis show that United State has maximum number of total minor and major accident in comparison to other country.



2. Aviation condition after accidents.

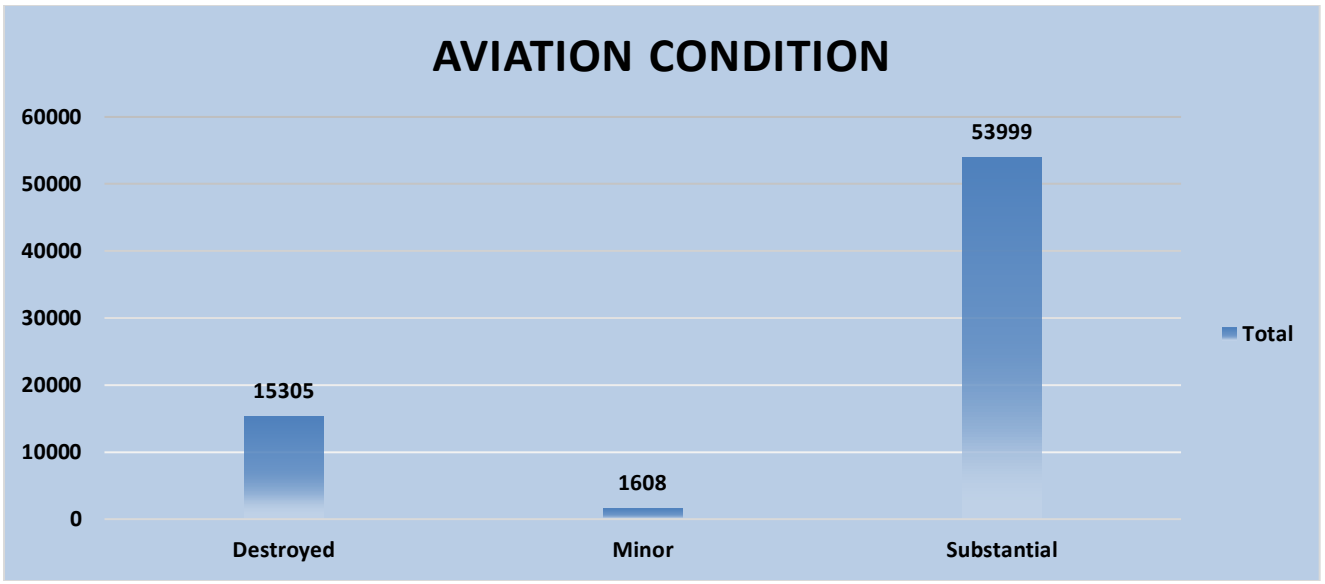
a) **Introduction:** This analysis show that what was the condition of aviation/flight after the accidents.

b) Specific Requirements/Functions and Formulas:

i. Pivot table of flight condition
ii. Pivot table of total accident
iii. Column Chart

c) Analysis Results:

• Analysis show that 15305 flight has been destroyed and 1608 has minor damage and 5399 has substantial.



3. **Total accidents according to types of engine.**

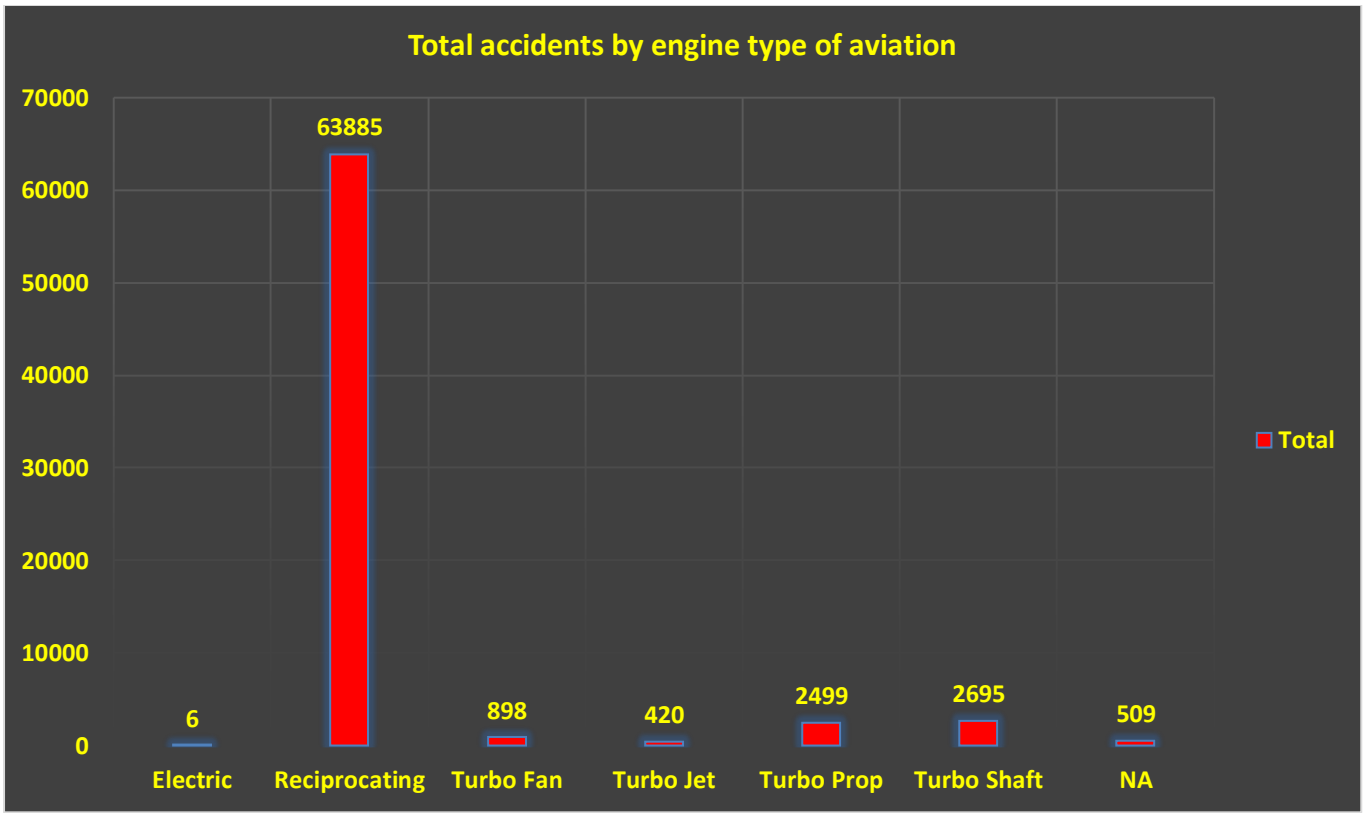
a) **Introduction:** This analysis show that which type of engine get maximum number of accidents and vice versa

b) Specific Requirements/Functions and Formulas:

i. Pivot table of engine type
ii. Pivot table of total accident
iii. Column Chart

c) Analysis Results:

• Analysis show that maximum number of accidents happened with Reciprocating engine type i.e, 63885 and minimum number of accidents happened with Electric engine type i.e 6.



4. Country-wise total number of accidents of the different purpose of flights.

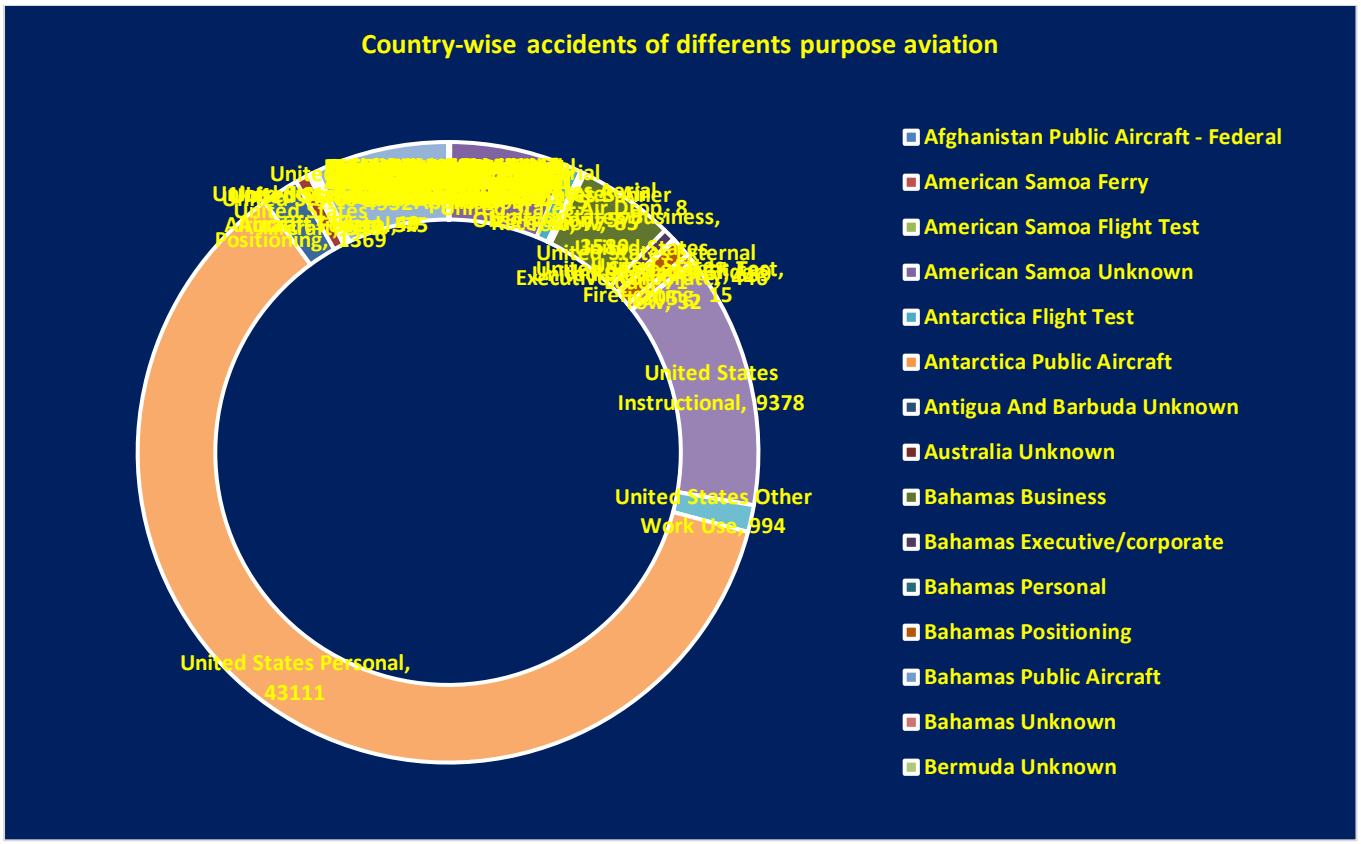
a) **Introduction:** This analysis show that country wise total accidents of different purpose of flights.

b) Specific Requirements/Functions and Formulas:

i. Pivot table of Country.
ii. Pivot table of Purpose of flights.
iii. Pivot table of total accidents.
iv. Pie Chart.

c) Analysis Results:

• Analysis show that United States had maximum number of accidents with Personal purpose flight i.e 43111 accidents.



5. **Total number of accidents segregated by years and months.**

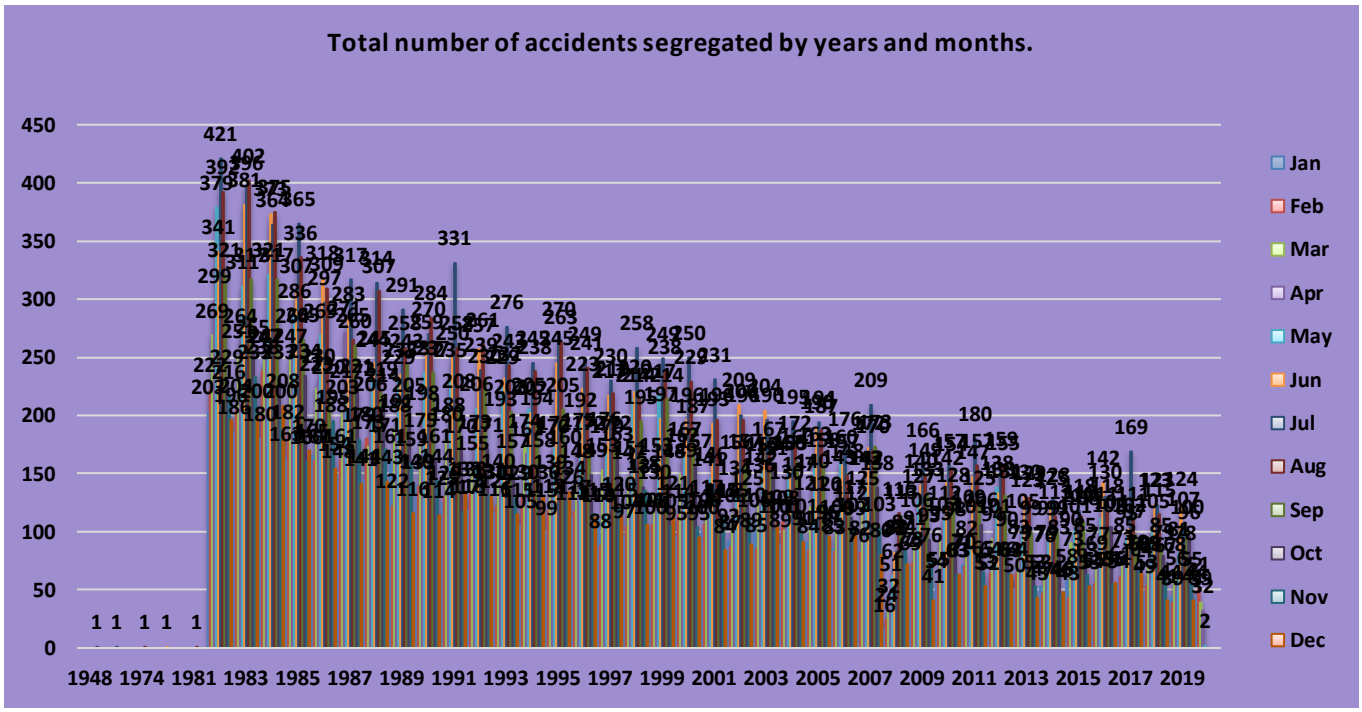
a) **Introduction:** This analysis show that total number of accidents every year each months happened.

b) Specific Requirements/Functions and Formulas:

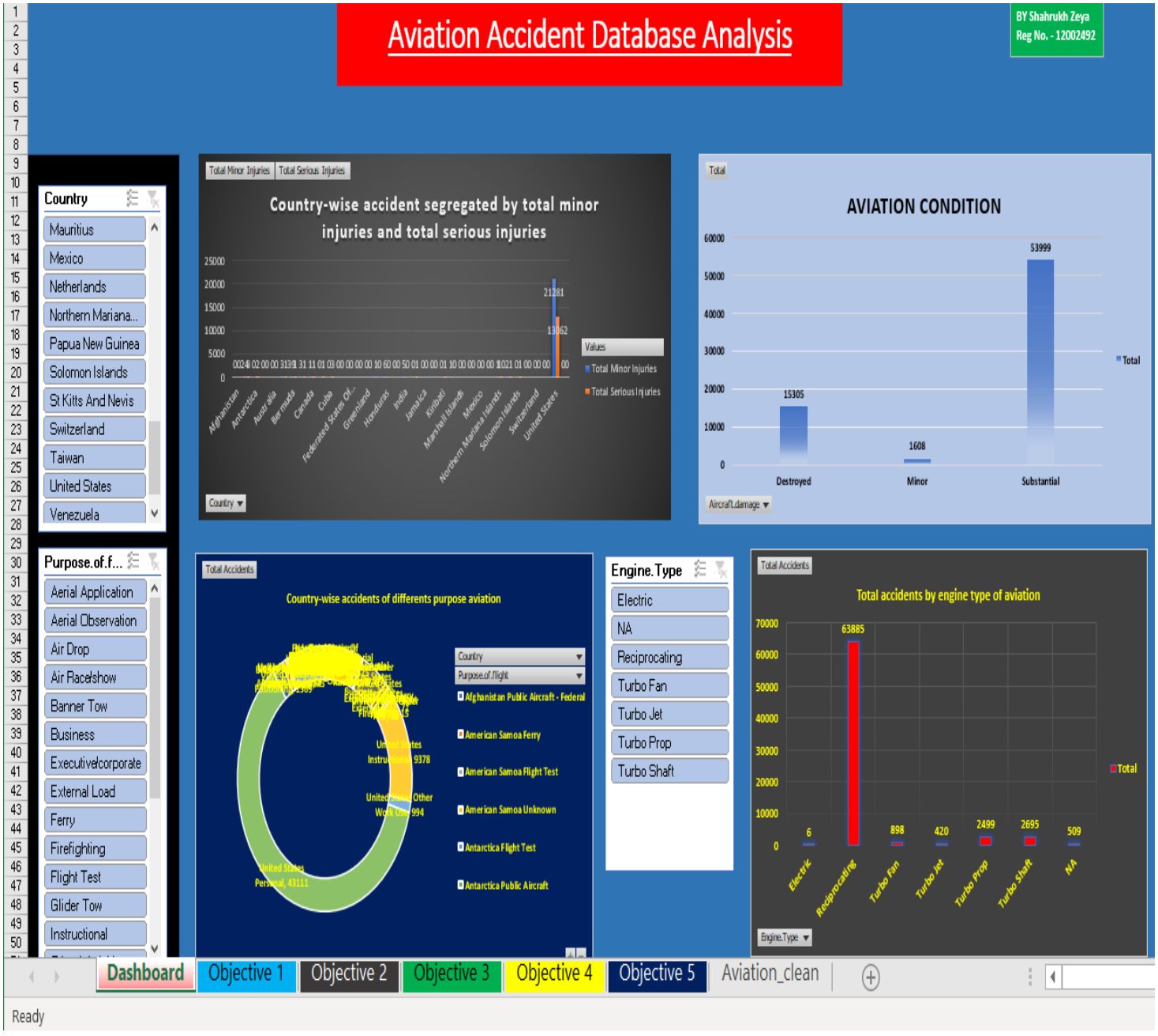
i. Pivot table of Year.
ii. Pivot table of Months.
iii. Pivot table of total accidents.
iv. Column Chart.

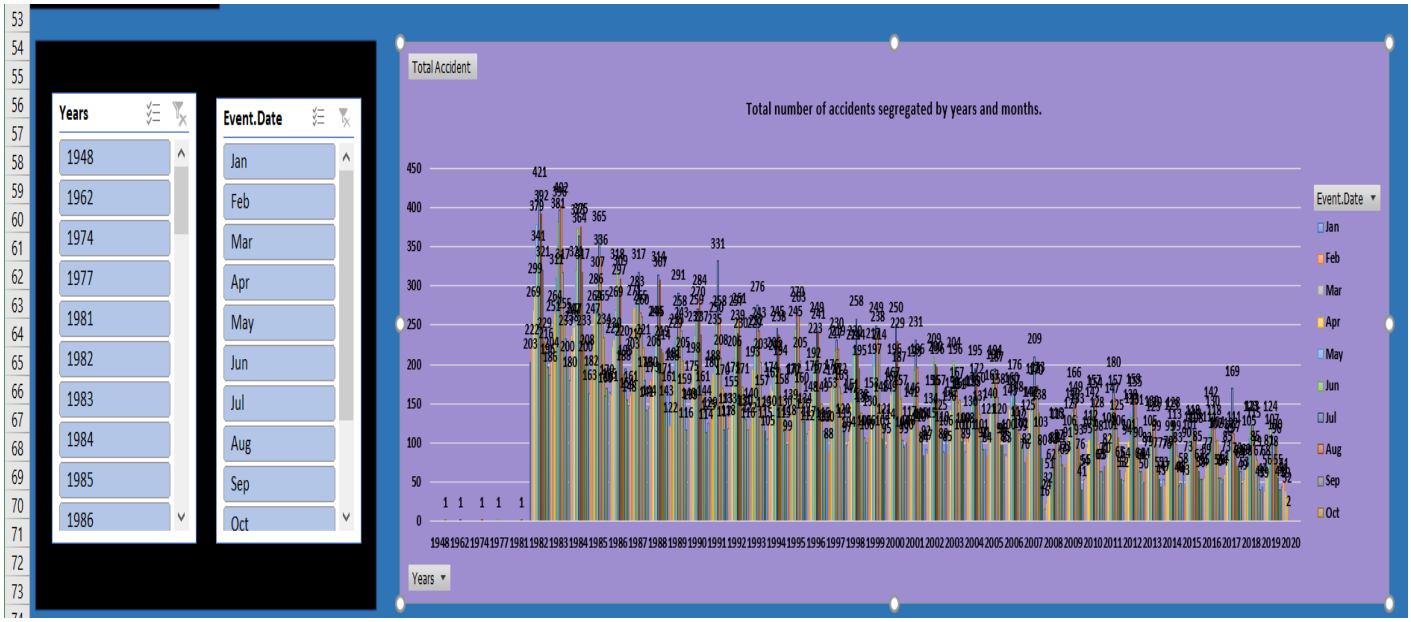
c) Analysis Results:

• Analysis show that July month had maximum number of accident in 1982 and there very low accident between 1948 to 1981.


Total number of accidents segregated by years and months.
450
400
350
300
250
200
150
100
50
0
1948 1974 1981 1983 1985 1987 1989 1991 1993 1995 1997 1999 2001 2003 2005 2007 2009 2011 2013 2015 2017 2019
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec

DASHBOARD


Aviation Accident Database Analysis
BY Shahrukh Zeya
Reg No. - 12002492
Country
Mauritius
Mexico
Netherlands
Northern Mariana...
Papua New Guinea
Solomon Islands
St Kitts And Nevis
Switzerland
Taiwan
United States
Venezuela
Purpose.of.f...
Aerial Application
Aerial Observation
Air Drop
Air Race/show
Banner Tow
Business
Executive/corporate
External Load
Ferry
Firefighting
Flight Test
Glider Tow
Instructional
Total Minor Injuries
Total Serious Injuries
Country-wise accident segregated by total minor injuries and total serious injuries
Values
Total Minor Injuries
Total Serious Injuries
Country
Total
AVIATION CONDITION
Destroyed
Minor
Substantial
Aircraft.damage
Total Accidents
Country-wise accidents of differents purpose aviation
Country
Purpose.of.flight
Afghanistan Public Aircraft - Federal
American Samoa Ferry
American Samoa Flight Test
American Samoa Unknown
Antarctica Flight Test
Antarctica Public Aircraft
Engine.Type
Electric
NA
Reciprocating
Turbo Fan
Turbo Jet
Turbo Prop
Turbo Shaft
Total accidents by engine type of aviation
Engine.Type
Dashboard
Objective 1
Objective 2
Objective 3
Objective 4
Objective 5
Aviation_clean
Ready



BIBLIOGRAPHY

1. www.kaggle.com
2. www.youtube.com
3. www.google.com
4. www.geeksforgeek.org
5. www.guru99.com