

Annexure-I (a)

Data Science Project Report

A report

Submitted in partial fulfilment of the requirements for the award of degree of

US Highway Railroad Crossing Accidents

Submitted to

ASHU MAM

LOVELY PROFESSIONAL UNIVERSITY

PHAGWARA, PUNJAB



SUBMITTED BY

Name of student: Kanigolla Naga Venkata Bala Likhith

Registration Number: 11903700

Subject Code : INT 217

Section: KM003

Annexure-I (b): Student Declaration

To whom so ever it may concern

I, **Kanigolla Naga Venkata Bala Likhith ,Registration, 11903700** , hereby declare that the work done by me on “**Data Science Report on US Highway Railroad Crossing Accidents**” and Submitted to **Ashu, Professor**, Lovely professional University, Phagwara, Punjab, is a record of original work for the partial fulfilment of the requirements for the award of the degree, **ETP Report**

Kanigolla Naga Venkata Bala Likhith (11903700)

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.

I consider it as great privilege to have esteemed Lecturer Mrs Ashu 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 such a great learning experience.

Kanigolla Naga Venkata Bala Likhith
11903700
KM003
Lovely Professional University

Table of Contents

S.no	Table of Contents	Page No
1	Introduction of the Data	5
2	Scope of the Analysis	6
3	Existing System	8
4	Analysis on Data Set	9
5	List of Analysis with results, Future Predictions	17
6	References	18

INTRODUCTION ABOUT THE DATA

This data is taken from US department of Transportation. The data is provided by FRA Office of Railroad Safety and dataset owner is Jared McCulloch. The dataset cover accidents from 1st January 1975 to 28 February 2021. It has 239487 rows and 141 columns. This data is extracted from the accident report form. [Form57](#) The features consist of Geography, Time Frame, Type of Crossing, Type of Accident, Type of Vehicle, Position, Type of Highway User, Type of Equipment, and/or Highway User Action.

Highway user involved:

Type

- A. Auto
- B. Truck
- C. Truck-trailer
- D. Pick-up truck
- E. Van
- F. Bus
- G. School bus
- H. Motorcycle
- J. Other motor vehicle
- K. Pedestrian
- M. Other (specify)

Scope of the Analysis and Other Information About the Data



The United States Department of Transportation is a federal Cabinet department of the U.S. government concerned with transportation. It was established by the Department of Transportation Act of Congress on October 15, 1966 and began operation on April 1, 1967. The secretary of transportation is the head of DOT.

The department's mission is "to develop and coordinate policies that will provide an efficient and economical national transportation system, with due regard for need, the environment, and the national defence."

DOT employs almost 55,000 people across the country, in the Office of the Secretary of Transportation (OST) and its operating administrations and bureaus, each with its own management and organizational structure.

Other Information About the Data (Original Data) :

Data Last Updated on:	5 th December 2021
No of Rows:	241K Rows
No of Columns:	159 Columns
Publisher:	Federal Railroad Administration, US
Contact Name:	Office of Railroad Safety, US
Bureau Code:	Federal Railroad Administration, US
Program Code:	Office of Safety, US
Language:	English
Licence:	http://www.usa.gov/publicdomain/label/1.0/
Official Website:	https://railroads.dot.gov/accident-and-incident-reporting/highwayrail-grade-crossing-incidents/highwayrail-grade-crossing
Data Provided By:	FRA Office of Railroad Safety, US
Dataset Owner:	Jared McCulloch

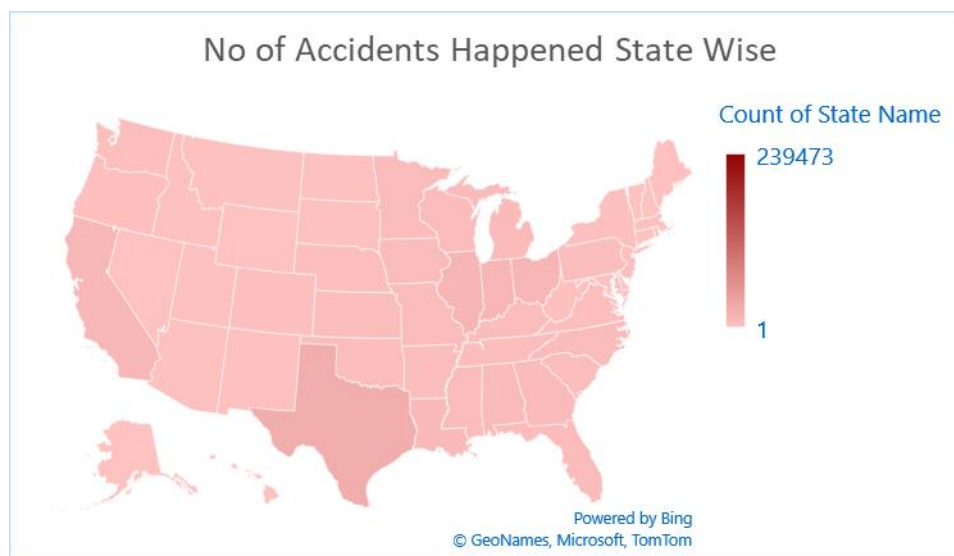
Existing System

Before existence of Data Science, analysing data used to be hectic task and existing system didn't used to analyses the data with perfection. Without existence of current cutting-edge technology of data science, we can get actionable insights in the dataset of the Indian Trade. Following are the benefits which weren't present in the existing system of data analysing:

1. Making Better Decision with The Help of Data
2. Directing actions based on trends- which later defines the goals required for profit.
3. Doing challenging stuffs with the help of prediction which is done by data.
4. Identifying various opportunities to increase the profit,
5. Making decision with Quantifiable, data driven evidence so that loss doesn't happens.
6. Testing the decisions taken by the data and watching and analysing the trend.

1. Data of the Accidents occurred State wise

- a) **Introduction:** The Analysis shows the number of accidents occurred in states of USA and returns us the data.
- b) **Specific Requirements/ Formulas and Functions:**
 - i. Taken the data of the state wise
 - ii. Added the Country Name to it
 - iii. Used the Count Function to get the data state wise
 - iv. Copied the pivot table to the side and made the map chart using that static table
 - v. Linked the data to the pivot table so that we can use it in the slicer and even if any data is updated then automatically it will get updated
- c) **Analysis Results:**
 - The report is used to monitor, evaluate, analyse, and determine the no of accidents happened month and year wise in the state map marking the results in the state from light colour to dark colour.
 - If we see the overall time starting from 1975 – 2021 we can see the top 10 states where the accidents occurred most is
 - TEXAS
 - ILLINOIS
 - OHIO
 - INDIANA
 - CALIFORNIA
 - LOUISIANA
 - MICHIGAN
 - GEORGIA
 - ALABAMA
 - FLORIDA
- d) **Visualization:**



2. Accidents occurred in different weather conditions:

a) Introduction: The data shows the sum of the accidents occurred in the different weather conditions.

The Weather Conditions are

- Clear
- Cloudy
- Fog
- Rain
- Sleet
- Snow

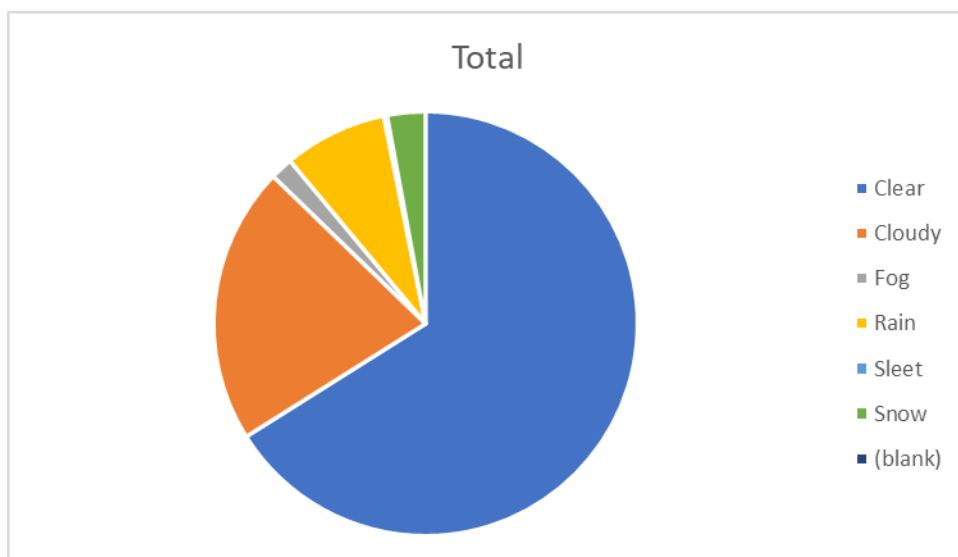
b) Specific Requirements/ Formulas and Functions:

- i. Pivot Table of the Weather Conditions
- ii. With the help of the data the Pie Chart is Plotted

c) Analysis Results

- We have different weather conditions in different seasons in predicting the future analysis it is good if we have the data in which we are having high accidents in which weather
- Most of the Accidents are occurring in the Clear weather conditions

d) Visualization



3. Total Number of Accidents occurred Year, Month, Day Wise :

a) **Introduction:** The analysis shows the data of the accidents occurred Year, Month, Day wise. We can see the ups and downs in the line chart so that can be visualized when the accidents occurred most.

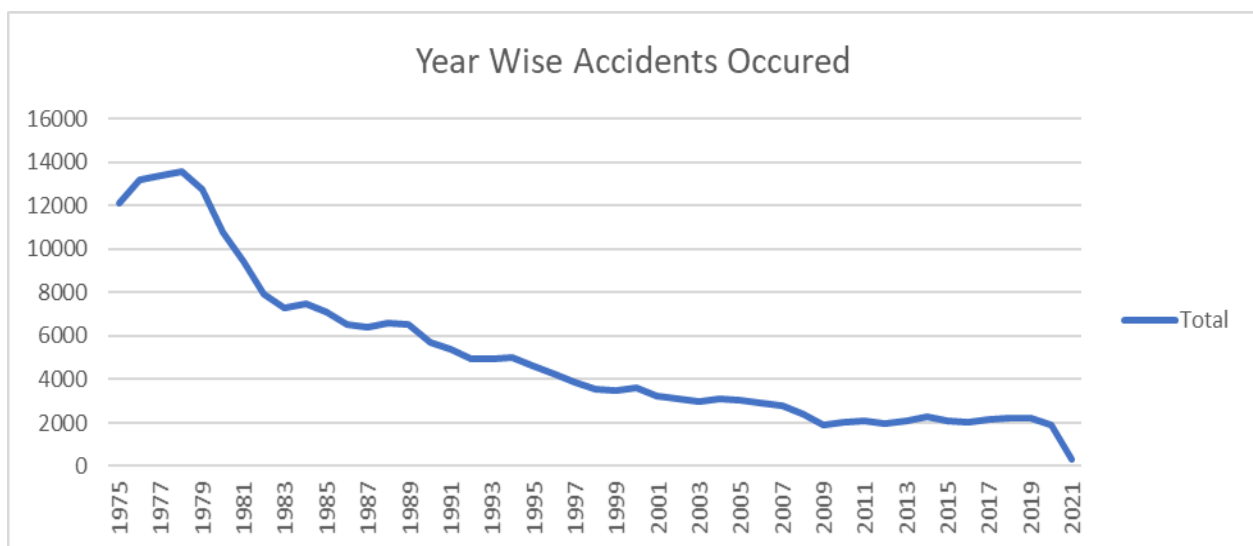
b) **Specific Requirements / Formulas and Functions:**

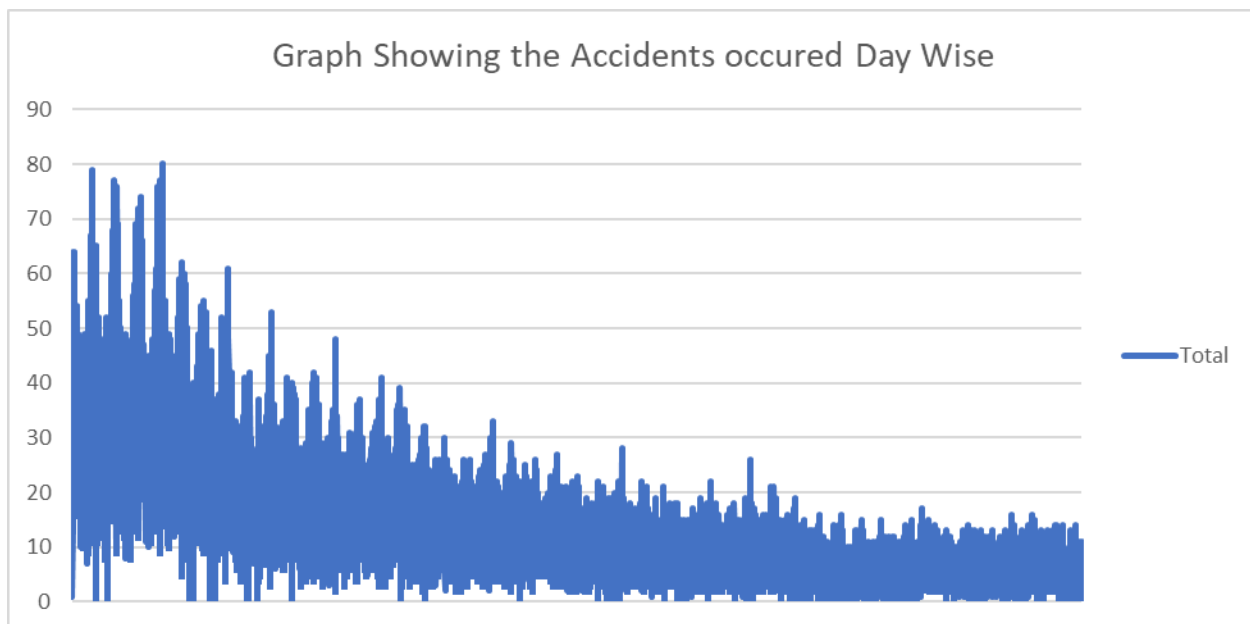
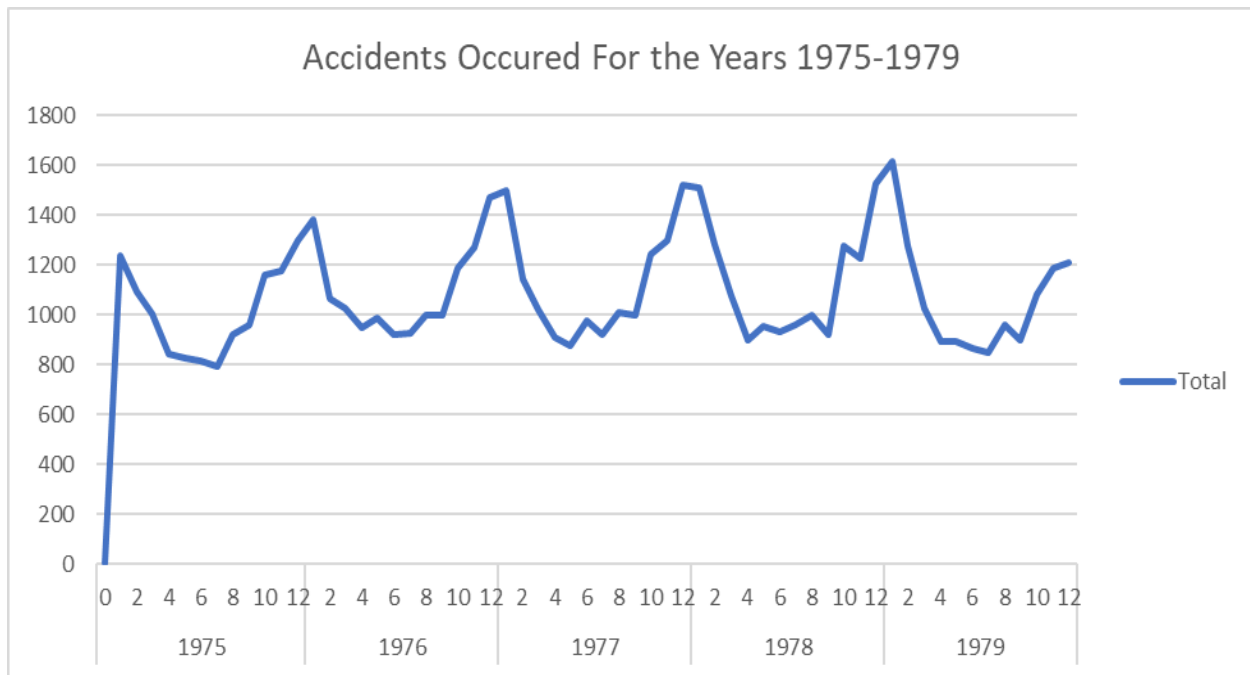
- i. Pivot table with the data of the years, month, and days
- ii. Added the Years in the Columns first and then followed by months, followed by days.
- iii. In this case we will get a classified data of the accidents occurred in the year wise and then in the months of that specific year and then days in that specific month
- iv. Making the line chart from the data we got from that table

c) **Analysis Results:**

- The Most accidents occurred in the years from 1975 – 1979
- After 1979 there is a high volume of decrease in the accidents till 1983
- After 1983 there are some ups and downs in the upcoming years till now
- The Highest Number of the Accidents occurred in 1978 with 13,556.
- If we consider the highest number of accidents occurred in a single month is 1,617 in the month of Jan in 1979

d) **Visualization:**



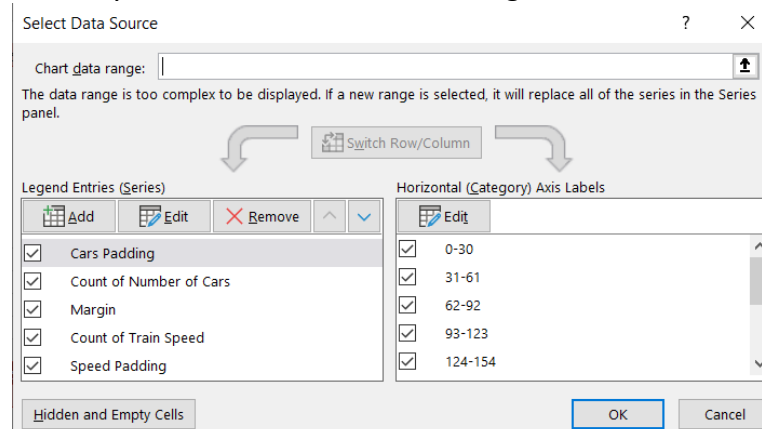


4. Total Accidents Occurred According to the Speed of the train and Cars count of the train:

- a) **Introduction:** A Butterfly chart has been plotted by considering the no of cars having for the train and train speed.
- b) **Specific Requirements / Formulas and Functions:**
- Created a pivot table with the no of cars count
 - Used the Count function to get the data of how many trains are in total
 - Cloned the same data and converted them into percentages
 - Created another pivot table just like the above one but with the data of train speed
 - Grouped the numbers in both pivot tables by 30
 - Cloned the information from the pivot tables and made the static table
 - Removed the percentage we got from the tables and kept as padding
 - Same done with another table
 - Here used VLOOKUP Function to get the exact things from the pivot table. If we have selected slicer and then we got some blank value, then the cell moves up as we cloned the data, we will get the wrong values
 - Also used IFERROR function because if we haven't got any value, it returns us error this function checks for the value and returns 0 if the value is not available in the pivot table
Example: =IFERROR((VLOOKUP(N27,\$N\$3:\$P\$13,2,FALSE)),0)
 - Here the static values are the range of the table and the padding. Final output looks like below

Row Labels	Count of Number of Ca	Count of Train Speed	Cars Padding	Speed Padding	Margin
0-30	49.21%	66.61%	50.79%	49.21%	50%
31-61	18.39%	30.73%	81.61%	18.39%	50%
62-92	16.86%	2.65%	83.14%	16.86%	50%
93-123	12.17%	0.01%	87.83%	12.17%	50%
124-154	2.61%	0.00%	97.39%	2.61%	50%
155-185	0.66%	0.00%	99.34%	0.66%	50%
186-216	0.09%	0.00%	99.91%	0.09%	50%
217-247	0.02%	0.00%	99.98%	0.02%	50%
248-278	0.00%	0.00%	100.00%	0.00%	50%
279-309	0.00%	0.00%	100.00%	0.00%	50%

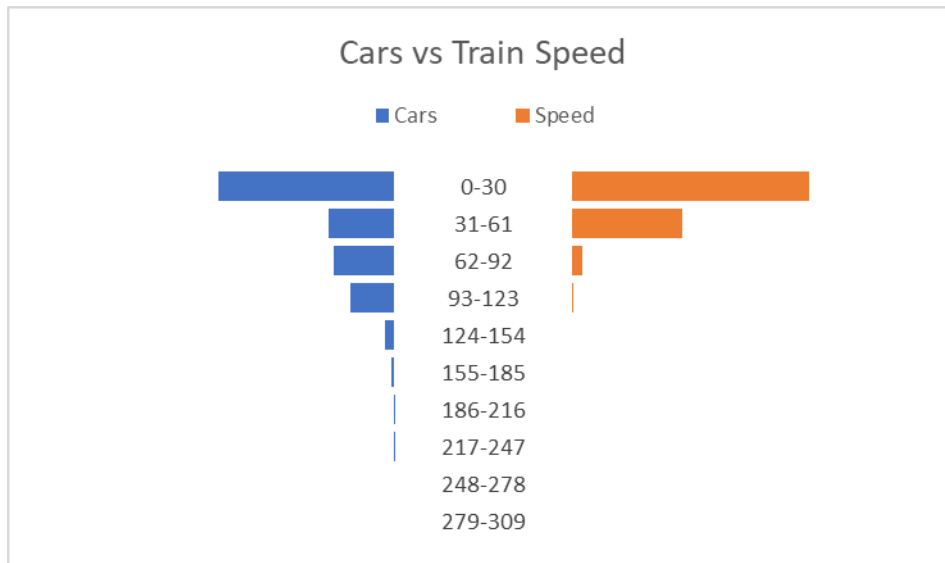
- xii. Made the stacked bar chart and hidden the values of the padding to get a perfect butterfly chart. The order of the arrangement can be seen below



c) Analysis Results:

- The trains having 0-30 cars and train travelling in 0-30 speed are involved in the accidents most

d) Visualization:



5. What are the Top Maintenance Roads Where Accidents occurred most:

a) **Introduction:** The railroads can be created by a company and might be maintained by the same company or the different one.

b) **Specific Requirements / Formulas and Functions:**

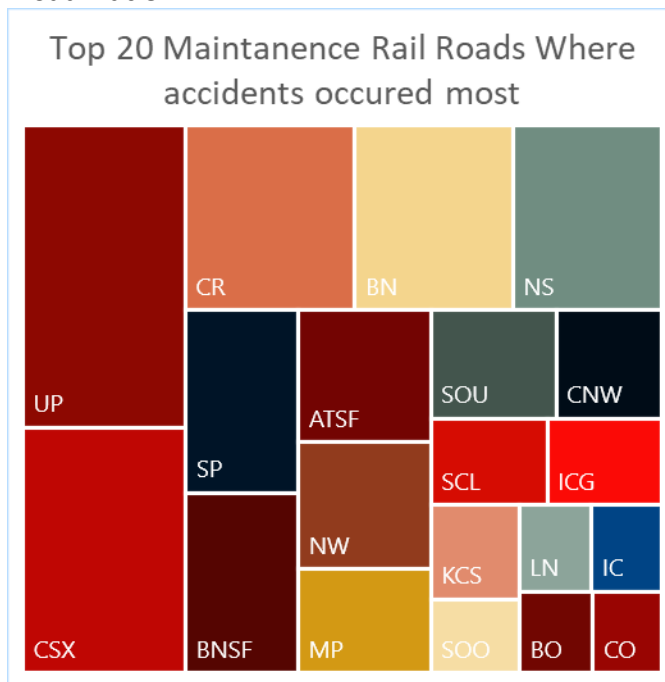
- Created a pivot table with the codes of the maintenance stations and sorted the top 20 where the most accidents occurred
- Generated the tree map from the obtained data.

c) **Analysis Results:**

If we considered the all the data of all the years, then the top is:

- | | |
|--------|-------|
| • UP | • SOU |
| • CSX | • CNW |
| • CR | • SCL |
| • BN | • ICG |
| • NS | • KCS |
| • SP | • SOO |
| • BNSF | • LN |
| • ATSF | • IC |
| • NW | • BO |
| • MP | • CO |

d) **Visualization:**



6. Total No of People Injured and Killed :

a) **Introduction:** A pie chart created from the sum of total deaths and total number of people injured

b) **Specific Requirements / Formulas and Functions:**

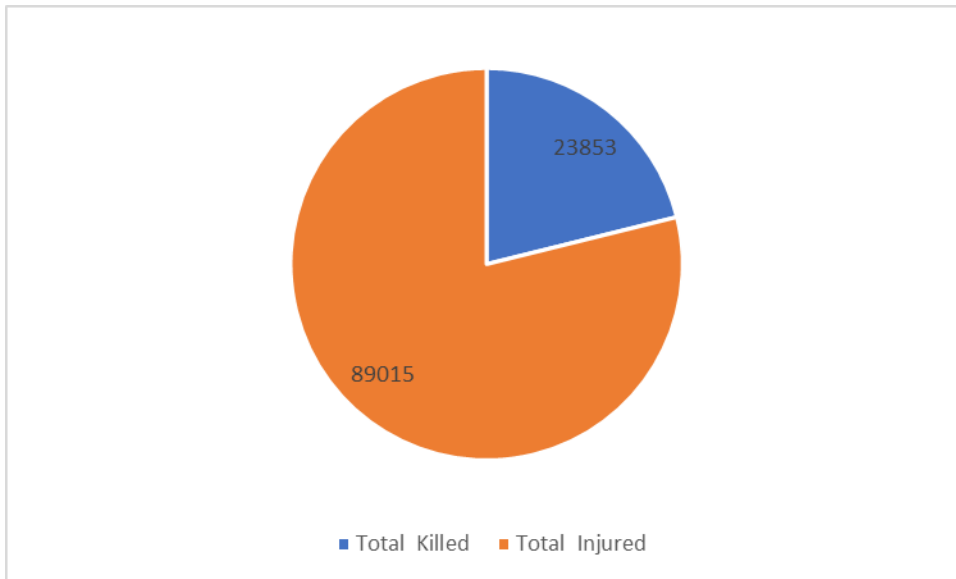
- In the Data set there are different categories
- Added the three columns of injured people to one and added three columns of the killed people data to one and made the pie chart from it

c) **Analysis Results:**

Total People got Injured in all years : 89015

Total People got killed in all years : 23853

d) **Visualization:**



Complete Analysis and Future Predictions

- Texas is the state where the most accidents occurred
- Highest Number of Accidents Occurred in the year 1979
- The Most accidents occurred in the years from 1975 – 1979
- After 1979 there is a high volume of decrease in the accidents till 1983
- After 1983 there are some ups and downs in the upcoming years till now
- The Highest Number of the Accidents occurred in 1978 with 13,556.
- If we consider the highest number of accidents occurred in a single month is 1,617 in the month of Jan in 1979
- The maintain stations of UP and CSX are having most accidents
- The Accidents are more in the month of March- April and in November – December
- The Average Train Speed is 25
- The highest Damage Cost is 1000000\$
- In an average case 12 people are getting killed and 100 people are getting killed
- Most Accidents occurred while the users moving over crossing
- The accidents are occurring in the 50F - 56F temperature

REFERENCES

<https://data.transportation.gov/Railroads/Highway-Rail-Grade-Crossing-Accident-Data/7wn6-i5b9>

<https://www.slideteam.net/blog/how-to-create-a-butterfly-chart-tornado-chart-in-powerpoint>

<https://excel.officetuts.net/examples/convert-a-pivot-table-to-table/>

<https://support.microsoft.com/en-us/office/tutorial-create-map-based-power-view-reports-62ecd29a-91b8-4e5c-b3b7-4aef9f7ef962>

<https://www.kaggle.com/yogidsba/us-highway-railgrade-crossing-accident>