28th JULY 2019

CSCI6612- Visual Analytics

PROJECT REPORT: TRENDS IN AVIATION ACCIDENTS

Team HOPE

VARUN GOPALAKRISHAN, B00823111
[varun.gopalakrishnan@dal.ca]

DEEPAN SHANKAR, B00804970
[deepan.shankar@dal.ca]

MAANASA TIRUVEEDHULA, B00817093
[mn233351@dal.ca]

JEEL GONDALIYA, B00819827
[Jl251073@dal.ca]

Contents

Abstract: 2
Project Explanation and its importance: 2
The Solution 2
Regression: 3
  Logistic Regression: 3
  Polynomial regression: 3
  Stepwise Regression: 3
  Ridge regression: 4
  Lasso regression: 4
  Elastic net regression: 4
  Linear regression: 4
Clustering 5
  Why k-means clustering? 5
The Visualization: 6
Conclusions and Future Work: 11
References: 12

Abstract:

As aircrafts have become the most common means of transportation, it must be made more reliable for people to travel without the fear of accidents. The increase in the number of accidents due to failure in the aircraft systems can be reduced if the incidents that are reported most often can be identified with their relevant features from the available dataset. The analysis of the information collected on the incidents of aircraft accidents helps in creating a model that can be used to predict the trends in accidents. The predicted trends can be used to understand accidents that occur frequently based on the climatic conditions or time of the year. Clustering the dataset based on the important features will help in identifying the categories under which each incident is grouped. This would improve the quality and safety of passengers traveling in aircrafts by identifying similarities with a known cluster. Further, by recognising the hotspots of accidents based on the incident location will help in finding regions where most incidents are reported. Hence, precautions can be taken while taking such air routes for travel. This analysis would help in avoiding accidents in the future to an extent by matching conditions that could result in an accident to that of parameters that are similar to an incident cluster.

Project Explanation and its importance:

The presents of extensive documentation of reported incidents in the field of aviation industry provides the opportunity to identify patterns and conditions under which accidents or system malfunctions take place in an aircraft. This project aims to identify patterns that could result in accidents which are a result of carelessness or improper organization in airlines to follow required system checks and suggest conditions which are safe for air travel. The analysis performed on this dataset would provide better understanding of the aviation accidents and help take precautions that would improve the safety of passengers during air travel.

The Solution:

As part of our proposed solution, a line graph is plotted to provide visualisation between the accident counts and the event date. The event date provides the date in which the incident occurred, and the accident count gives the number of accidents that were reported on that particular date. This graph provides an interactive interface for the users to select the year in which the users wish to see the accident rate. The line chart dynamically fetches the data for the particular year and displays the trends. The line chart also provides users with the ability to drill-down on a particular year to see the month and day in which the accident took place. This zooming feature provides better user experience by providing additional details on the accidents reported. Also, this helps in providing a consolidated view to the users by encompassing huge volume of data into a single line chart. The future trends in the accident rate was calculated by building a model using linear regression. As per the results, the accident rates are declining in the future.
The location of each incident reported is extracted from the aviation dataset to plot a heatmap that provides the hotspots of accidents on the map. This map provides a transition effect in which the users can visualize the change in the region of the hotspots as the trends in the accidents progresses with the years. This would provide the users with the accident-prone zones on a map.

K—means clustering is used to group similar incidents found in the dataset together. The clusters created from the algorithm is used to plot a bubble chart to visualize the similarity between the accidents. Based on the fatality rate, accident count, location etc. associated with a cluster, the future trips can be linked to a cluster for understanding the similarities. This would help in taking precautions to avoid known accidents. The size of the bubbles represents the number of fatalities in the accident and the colour of the bubble shows the country. The x and the y axis in the bubble chart represents the accident count and number of uninjured people.

Regression:

Linear regression models the relationship between two variables where one variable is an explanatory variable and the other is a dependant variable by fitting the linear equation to the observed data. If the approach is used for single explanatory variable, it is called simple linear regression and if the approach is for more explanatory variables then it is called multiple linear regression. The various correlated variables are predictions and are done on dependant variables which is called multivariate linear regression.[5]

Linear regression identifies statistic relationship but not the deterministic relationship. To determine the relationship between two variables, statistical relationship is inaccurate. The main idea of linear regression is to obtain a line that fits the given data in a best way.
It is one of the regression techniques which explores the relationship between dependent and independent variables. The advantage of using regression technique is that it indicates significant relationships between dependent and independent variables, also the strong impact of various independent variable on a particular dependent variable. This technique compares the effects of variables which are measured using different scales. Regression technique helps in choosing the best set of variables to evaluate and build the predictive models. There are various types of regression techniques such as linear regression, logistic regression, polynomial regression, step wise regression, ridge regression, lasso regression and elastic net regression.[7]

Logistic Regression:

Logistic regression is applied when the dependent variable of the data set is categorical. The data is first fitted into linear regression model which is later applied with logistic function to determine the predicted dependent variable which is categorical. For example, it can be used to predict whether an email is spam and whether there is malignant tumor in a person. There are various types of logistic regression such as binary logistic regression, multinomial logistic regression and ordinal logistic regression. The metrics used to evaluate logistic regression are sigmoid function, decision boundary and cost function.[12]

Polynomial regression:

Polynomial regression fits a relationship which is non-linear, between the value of ‘x’ and the mean value of ‘y’. Even though the polynomial regression fits the data to a non-linear model, it is linear if statistical estimation problem is considered, the regression function for the estimated unknown values is linear. The polynomial regression is also considered as the special case for multiple linear regression.[13]

Stepwise Regression:

Stepwise regression is a regression model which involves automatic selection of independent variables in a step-by-step iterative method. It is the combination of forward and backward selection techniques. The forward selection is modified in stepwise regression technique so that a variable is added after each step and all the candidate variables are checked to identify whether their level of significance has been reduced below the tolerance level. The insignificant variables found are removed from the model. Stepwise regression involves two methods for adding and removing the variables. To avoid getting into an infinite loop, the probability of cut-off for adding the variables should be less than the probability of cut-off for removing the variables.[14]

Ridge regression:

Ridge regression is a technique used to analyse the data which is affected by multicollinearity. In multicollinearity, the estimates of least squares are unbiased, and the variances may be large and far away from the actual value. The standard errors in the ridge regression can be reduced by adding the regression estimates with the degree of bias. Principal components regression is another type of biased regression technique, but the ridge regression technique is considered the best among two methods.[15]

Lasso regression:

Lasso (Least Absolute Shrinkage and Selection operator) regression utilizes the concept of shrinkage. The data points are shrunk towards the central point called mean. This regression type is well suited for the models that show high multicollinearity and also when there is a need to automate variable selection or eliminating parameter. It performs regularization which adds the penalty that is equal to the coefficients with the absolute value of the magnitude. The coefficients which become zero are eliminated from the model. If the coefficient value is close to zero, it indicates large penalties and is ideal to produce simple models. Ridge regression does not eliminate the coefficients which makes it easy for Lasso regression to interpret better than the ridge regression.[16]

Elastic net regression:

Elastic net regression is preferred when compared to lasso regression(L1) and ridge regression(L2) as it solves the limitations of both methods. The solution in elastic net regression is to combine the penalties of both lasso and ridge regression to obtain the best one out of two and also tries to minimize the loss function.[17]

Linear regression:

Linear regression is simple and is commonly used when compared with other regression models for predictive analysis. The metrics used for evaluating the model in linear regression are R squared value, correlation coefficient, null hypothesis and p-value.
R squared value is a statistical measure of checking the data which is closely fitted to the regression line.
R-squared = Explained variation / Total variation
The strength and direction of a linear relationship between two variables can be measured using the correlation coefficient. The coefficient values generally range in between -1.0 and 1.0. There occurs an error in correlation if the calculated number is greater than 1.0 or less than -1.0. If the value of correlation is -1.0 it indicates the perfect negative correlation and if the value is 1.0 it indicates the perfect positive correlation. There occurs no relationship between the two variables if there is a correlation of 0.0.

Null hypothesis is a type of hypothesis which proposes that there exists no statistical significance in a given set of observations. It shows no difference between the set of variables. Null hypothesis assumes to be true till the value is zero for statistical evidence.[6]
The p-value is a number which is interpreted between 0 and 1 using small p-value, large p-value and the value near to the cut offs. A small p-value which is typically less than 0.05, gives strong evidence to reject the null hypothesis. A large p-value which is typically more than 0.05, gives strong evidence to consider the null hypothesis.[6]

In this project, linear regression is used to predict the aviation accidents for upcoming years by utilizing the data of aviation accidents that occurred from the years 1948 to 2018. The linear regression model is evaluated using cross validation to determine the score of degree to predict for further years.

Clustering

Clustering is a Machine Learning technique, It creates the groups of data points based on similar traits of data points. These data points should be classified in such a way that all the similar data points should be in one clustering, Data points which are in different clusters should have dissimilarity in traits or properties. Clustering is unsupervised learning as it is helpful to find meaningful patterns.[2]

Clustering can be used to gain insight into our data-based on how many similarities have between data and the amount of data comes under a single cluster.

There are four types of clustering methods;

Density-Based Methods: This method creates a cluster of concentrated data points and considers empty as outliers.[4]

Hierarchical Based Methods: This method creates clusters which have predefined order.

Partitioning Methods: In this method, initially, all the data points are in a single cluster and then partitioned according to the distance from each other.[4]

Grid-based Methods: This method more focuses on the dense area value rather than the data points.

Clustering algorithms, which we have thought to try, are discussed below in brief:

K-Means: This algorithm is easy to use as well as it performs faster than the other algorithms. Moreover, the clusters are assumed to be even size, so this algorithm suits for visualization purpose. However, It leads to less accuracy as the small change in the data leads to high variance in the result. [3]

DB-SCAN: It is a density-based algorithm for clustering. It is not very sensitive to outliers, unlike K-means. However, DB SCAN does not work well with high dimensional data. [3]

However, we chose partitioning methods as we had the number of clusters. K-means algorithm is used for clustering where k is the number of clusters.

Why k-means clustering?

We wanted to cluster for our dataset to find a useful pattern in the datasets for each unique country. So, we are summing up all the injury count for each unique country and also the maximum occurrence of each categorical columns for all the countries. After that, we are applying a clustering algorithm on a dataset.

We chose the K-Means algorithm for clustering as we previously had the number of clusters. The clustering algorithm is applied to all the categorical columns as well as numerical columns. In our case, the number of the cluster we want is 10.

This figure given below explains the process flow of the k-means clustering algorithm. Initially, the number of clusters is given, so If the number of clusters is k, then k centroid will be chosen randomly.

Iteratively, for each data point, the nearest centroid is selected until none of the assignment of the cluster's changes. In the end, all clusters are created based on the distance between data points and centroid. But if we add data then clusters might be changes. Hence, the K-Means algorithm is very sensitive to outliers but as it is very easy to understand and to implement, this approach quite popular in data science to classification or to find useful patterns.

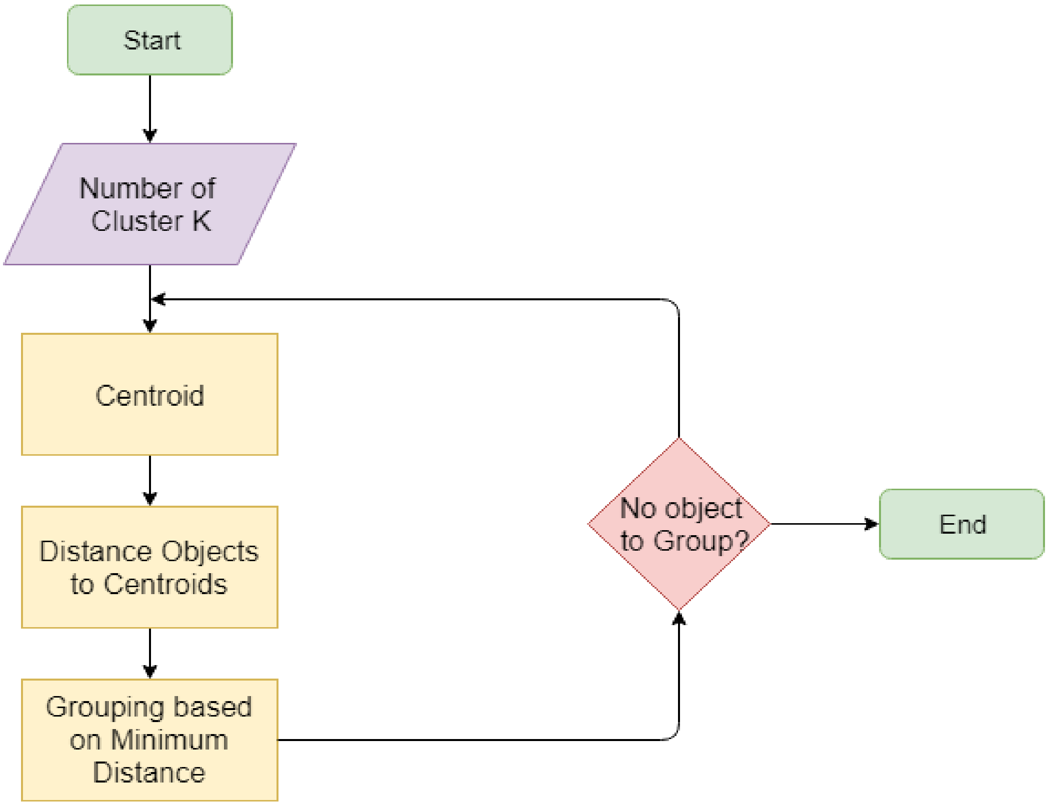

Fig 1. Flow chart of K-Means Clustering [1]

Initially, we have selected important attributes and got rid of all the irrelevant attributes. So we had these attributes after eliminating unnecessary attribute; latitude, longitude, phase of flight, country, weather condition, total minor injuries, total major injuries, total fatalities etc. get_dummies() method is used to convert all categorical columns into binary digits and we have also normalized the data before clustering.

After given the number of clusters, the K-Means algorithm is applied to the subset of important features and we saved the cluster labels in the CSV file.

The Visualization:

The visualization for the airline accidents from 1982 – 2017 is done using 3 different graphs, each depicting different visualization of attributes and helps the user to easily visualize the data. We have provided each tab for different visualization where the user can select the tab of visualization they are interested. The three visualization are multiple line chart for trend, Heat Map and bubble chart using clustering and grouping.

The first chart is a multiple line chart for trend which depicts the number of accidents for each year with its date and its corresponding number of injuries. Each line depicts the number of injuries, number of uninjured, number of fatalities. The user is able to identify the trend in the airline accidents which took place for almost 30 years and is able to identify the pattern in which the accident rates change. The user is also able to identify whether the accidents rate increase or decrease over the period, thus analysing the trend, gives the user a visual comparison with each year. The user also can identify the dates when the number of accidents has increased and also have the details which year has highest accidents and can analyse based on this information later. The user can also select the year which is provided by a dropdown and by selecting the year, the trend is plotted for only the selected year. This provides the user to identify the trend on the specific year rather than having a complicated trend chart for all years. This also provides for the user to find the number of accidents during a particular date of the year by selecting the year and identifying the date which provides a simpler approach to find number of accidents for the specific date.

The chart also plots the number of accidents for future years (2019-2021) which is provided by linear regression which predicts the number of accidents for future years based on certain attributes and parameters. This provides the user to get an overall estimate for the number of accidents in the future. Visualizing this data in a line chart helps the user to identify and compare the accident rates visually in a more efficient way and helps the users to have an overall trend for the past years. This chart also provides to investigate insights of each month of each year by zooming the chart which give the drilldown visualization for each day and/or each month of the year. This zoom functionality gives the users to quickly find the data value for a specific date by zooming into the chart.

Other features provided for this chart are the scrolling of the zoomed area in the display at the bottom of the chart which displays the zoomed portion in the original chart above and a tooltip on the line provides the values for the particular mouse-over date of the selected year for all the lines. This provides a quick look at the value at any instant date on the line chart which is useful to know the injuries count by hovering on the line for each day. Thus, the line trend chart provides the user to quickly visualize the accident trend for the past and future years and also provides an efficient way to compare the accident rates.

We initially thought to implement this data using bar chart, but we later decided that bar chart has one bar for each date and pre-processing to combine the data into monthly or weekly helps to visualize the data better. This is considered as future work. Thus, a line chart with zoom would be a good option for the users to visualize these data for finding the pattern in the trend and helps the users to quickly analyse the data and work on it.

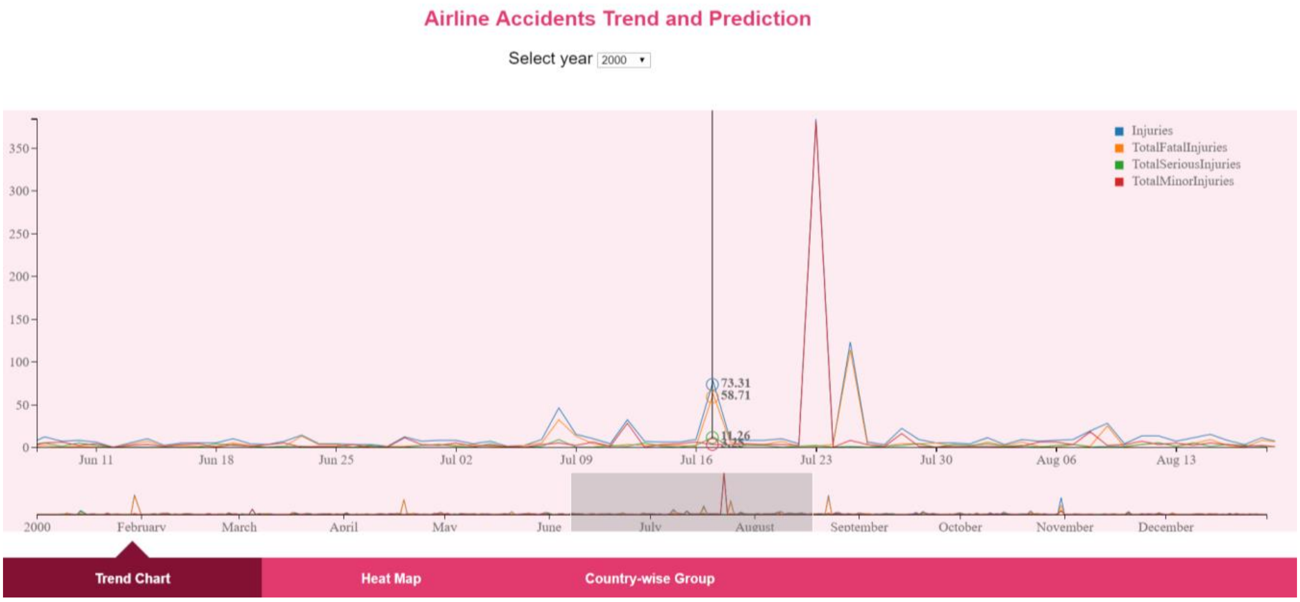

Fig 2. Line Chart for Trend Analysis

The second graph is a geographic heat map visualization of different countries on the geographic map for past years (1948 – 2018). This graph provides the user to identify the area which are more concentrated on the accidents depicted in a heat map visualization. This provides insights of the geographic location where there is large concentration on accidents and also provides areas where accidents prevail less likely. The heat map provides areas of accidents for each year where accidents have occurred and plots the areas on the geographic map where accidents prevail a lot and highlights those areas as more concentrated and more accidents have taken place in those regions during a particular year. This map provides a visualization for the users to have a good understanding on the areas or regions where accidents have taken place and also the users could identify the high and low concentrated areas on the map by looking at this visualization.

Since we had the longitude and latitude information, we thought to locate each accident on the geographic map with each point locating to a particular area. But there are many accidents within each area, plotting the data in map would not satisfy our solution. Thus, we had an idea to implement it using heat map which describes the concentration.

This visualization plots the data as per the number of accidents that have taken place during a particular year across the world and highlights by areas of dark colours for the areas where the number of accidents is more than the areas of less accidents. This also helps the users to find out the area where it is safe place to fly an airline by identifying the less concentrated areas where accidents have happened less frequently. Thus, providing the airline team the areas which are safe to fly. This visualization provides the heat map concentration only for a particular year but also provides a timer of 2 seconds where the map data changes for other years based on the timer. Thus, the user is able to visualize the heat map data for all the years based on timer and analysing data for a particular year.

By analysing the visualization, the user is able to identify the pattern in which there is increase in accidents in a particular region across the geographic area over the period and provides the user to analyse the trend in areas where accident rates have increased or decreased. The user is also able to identify the risk factor in flying the airline through the areas of highly concentrated accidents and also could analyse the airlines to find the reason behind the areas of more number of accidents by visualizing the data in a geographic map.

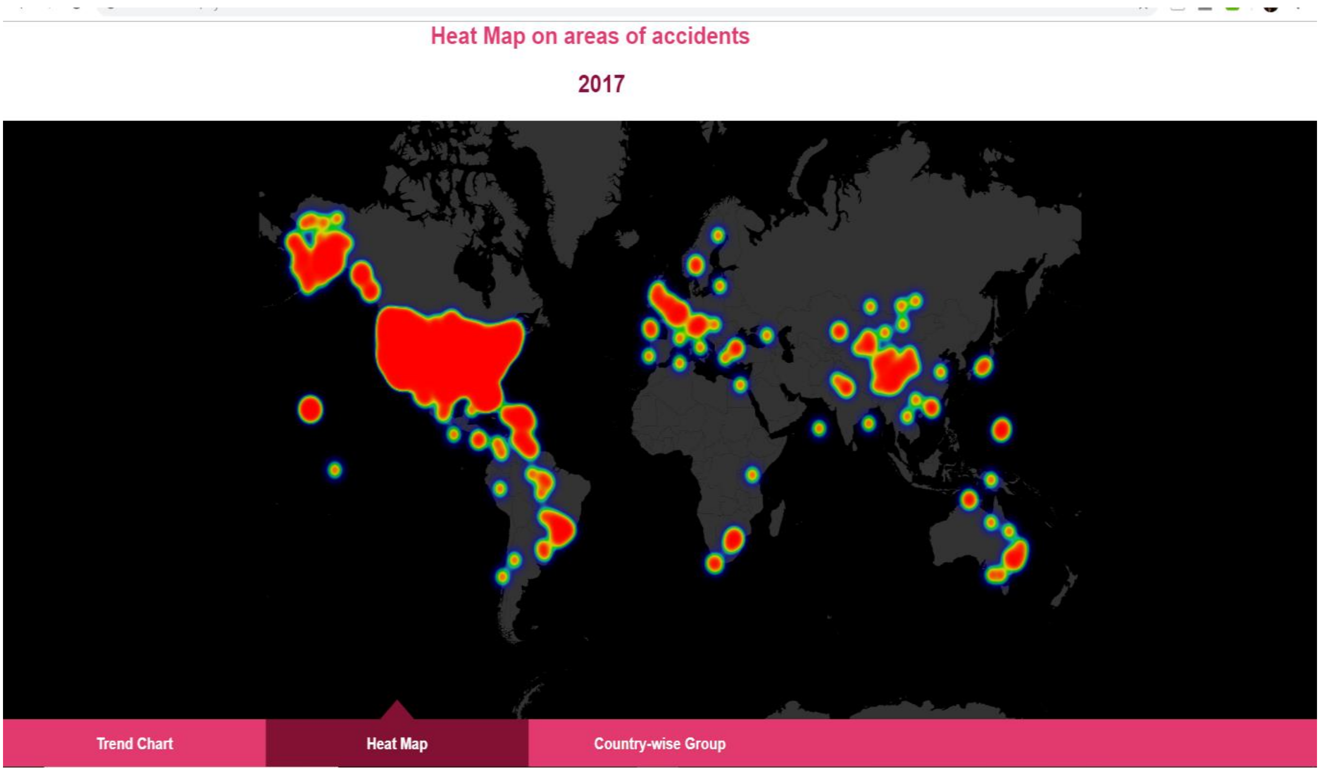

Fig 3. Heat Map for Trend Analysis

The third graph is a bubble chart which depicts the data by clustering and forming groups and identifying each country to from a group. This visualization provides the user to find the countries of a particular group and identifies the clustering of different countries. The clustering of countries is done by using k-means algorithm and the initial cluster size is defined as 10. Thus, the data is clustered into 10 clusters based on the attribute's countries, phase of flight, climatic conditions, latitude, longitude, make and model of flight, number of injuries, number of minor and major injuries, number of fatalities. After the clustering and forming clusters, these data is grouped based on the attributes and the data is plotted as a bubble chart.

The data for bubble chart is provided by filtering out the countries, number of total injuries, number of major and minor injuries and number of fatalities and the cluster number. The bubble chart is plotted with these values with total number of major and minor injuries on x-axis and total number of all injuries on y-axis. Each bubble depicts the different countries and the size of each bubble depicts the number of fatalities. Each cluster is identified by the different colours where each colour depicts a cluster.

The user is able to identify the groupings of each country in cluster and is able to visualize the data based on the clustering and helps the user to identify the countries which follows the pattern by finding the countries in the specific clusters by identifying the colour of each cluster. It also provides the user to categorize the countries based on different attributes like number of injuries and number of fatalities and is able to identify the similar countries which follows this cluster. Hence providing the user to distinguish the countries based on the attributes. It also provides the user to find the number of fatalities by looking at the size of the bubble to find the number of fatalities of all the countries and compare them by looking at their size. This gives the user to visually identify the country having the highest number of fatality count and the country with lowest count.

We have provided user interaction on this chart by adding tooltip on each bubble which displays the country name with the colour of its cluster as the colour of tooltip. It provides the user to hover on the bubbles and identify the country of their interest. The legend provided on this chart displays each cluster with different colours and hovering on the legends will highlight the bubbles of that cluster and the user is able to identify all the countries of that cluster. This provides the user to hover on the legends and find out the similar countries of that clusters. The user is also able to visualize the countries with low injuries and identifying those clusters. Thus, providing the users to identify the clusters with low number of injuries.

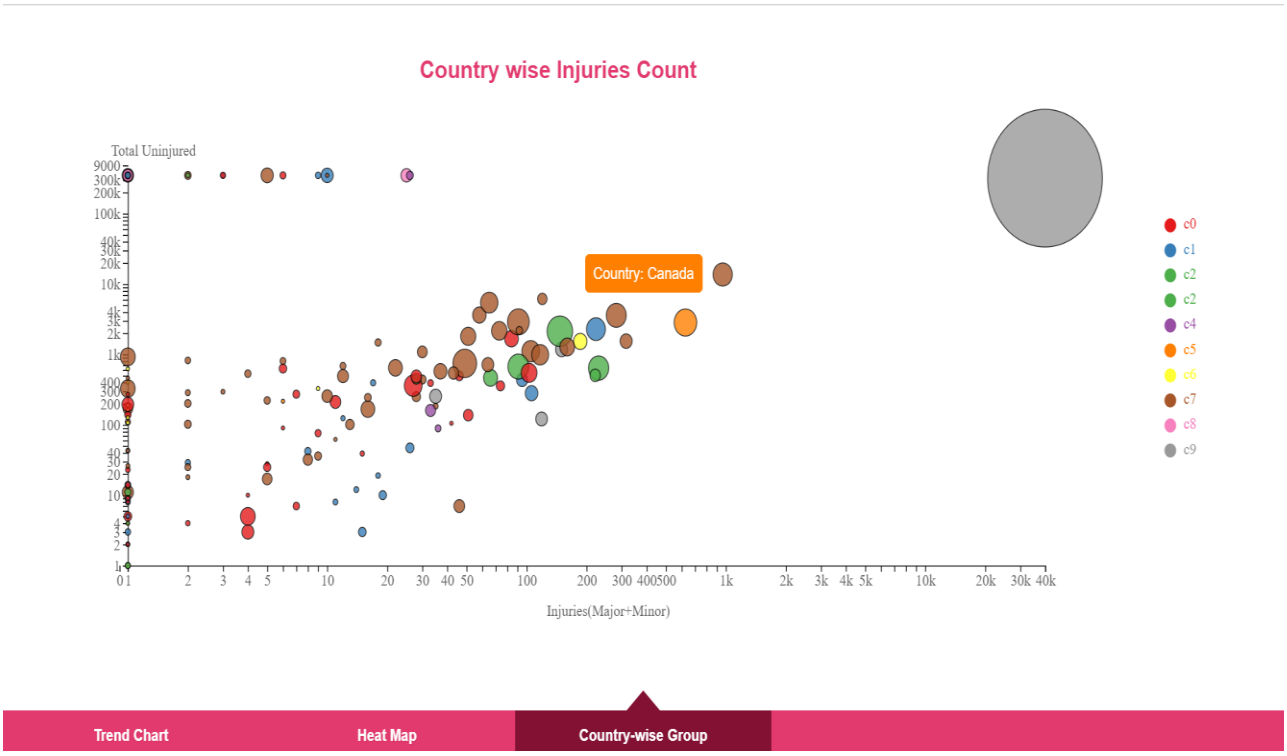

Fig 4. Bubble Chart Using Clustering

The evaluation is based on the time in rendering the graphs which nearly takes less than 1 second for all charts. A timer is provided for the heat map which changes the data of heat map every 2 seconds. It also provides smooth transition while hovering for tooltip and legends.

Conclusions and Future Work:

The application developed as part of this project provides users with the much-needed information about the trends and similarities present in aviation accidents. The presents of huge volume of data helps in performing an exploratory analysis on the incidents reported. Pre-processing the data before building a model to predict the trends in aviation accidents helped improve the model's results. As part of future work, other algorithms will be explored to perform regression on the extracted features to improve the trend analysis. The frontend visualization will be made mode interactive in a way that users will be able to compare the results from different machine learning algorithms.

The K-means clustering performed on the dataset to group similar entities together will be explored further by providing users with the option to choose the number of clusters formed. This option will be provided in the fronted for the users to visualise the bubble chart which would be generated as per the user input. Other clustering algorithms will be explored and implemented for the users to compare the results in the fronted visualization.

The Heat-map which provides the hotspots of accident on the map will be upgraded to have tooltips that would provide details about the particular incident on the map. The user will be provided with an option to pause the transition shown on the map to view the accident hotspot for a particular year. Since most hotspots are in and around the air route taken by the flights, it gives a false pretext that other area safer to fly. Hence, an efficient approach will be developed to highlight the frequent air routes on the heatmap so that the users will have the ability to compare accident hotspots based on the air routes taken by the aircraft.

References:

[1] *Cad.zju.edu.cn*, 2019. [Online]. Available: http://www.cad.zju.edu.cn/home/zhx/csmath/lib/exe/fetch.php?media=2011:presentation_ml_by_ibrar.pdf. [Accessed: 27- Jul- 2019].

[2] "The 5 Clustering Algorithms Data Scientists Need to Know", *Medium*, 2019. [Online]. Available: https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68. [Accessed: 27- Jul- 2019].

[3] "Understanding K-means Clustering in Machine Learning", *Medium*, 2019. [Online]. Available: https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1. [Accessed: 27- Jul- 2019].

[4] D. Science and A. clustering, "Clustering Introduction & different methods of clustering", *Analytics Vidhya*, 2019. [Online]. Available: https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/. [Accessed: 27- Jul- 2019].

[5]*Linear Regression*. Available: http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm.

[6]/@saishruthi.tn, "Linear Regression - Detailed View," *Medium*, 18-Jan-2019. Available: https://towardsdatascience.com/linear-regression-detailed-view-ea73175f6e8 6.

[7]S. Ray and Business Analytics, "7 Regression Types and Techniques in Data Science," *Analytics Vidhya*, 24-Jun-2019. Available: https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/.

[8]"What is Linear Regression?," *Statistics Solutions*. Available: https://www.statisticssolutions.com/what-is-linear-regression/.

[9] A. Ganti, "Correlation Coefficient Definition," *Investopedia*, 31-Mar-2019. Available: https://www.investopedia.com/terms/c/correlationcoefficient.asp.

[10]A. Hayes, "Null Hypothesis Definition," *Investopedia*, 25-Jun-2019. Available: https://www.investopedia.com/terms/n/null_hypothesis.asp.

[11]"What a p-Value Tells You about Statistical Data," *dummies*. Available: https://www.dummies.com/education/math/statistics/what-a-p-value-tells-you-about-statistical-data/.

[12] /@saishruthi.tn, "Logistic Regression - Detailed Overview," *Medium*, 18-Jan-2019. Available: https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc.

[13]/@lucamel7, "POLYNOMIAL LINEAR REGRESSION," *Medium*, 29-Aug-2018Available: https://medium.com/data-py-blog/polynomial-linear-regression-1f8117f6de2e.

[14]W. Kenton, "Stepwise Regression," *Investopedia*, 12-Mar-2019. Available: https://www.investopedia.com/terms/s/stepwise-regression.asp.

[15]*Ncss-wpengine.netdna-ssl.com*, 2019. Available: https://ncss-wpengine.netdna-ssl.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Ridge_Regression.pdf

[16]“Lasso Regression: Simple Definition,” *Statistics How To*, 14-Oct-2018. Available: https://www.statisticshowto.datasciencecentral.com/lasso-regression/.

[17]GeorgeOfTheRFGeorgeOfTheRF 1, SycoraxSycorax 45.4k1515 gold badges118118 silver badges216216 bronze badges, and Richard HardyRichard Hardy 29.1k66 gold badges4848 silver badges139139 bronze badges, “What is elastic net regularization, and how does it solve the drawbacks of Ridge (L^2) and Lasso (L^1)?,” *Cross Validated*, 01-May-1966. Available: https://stats.stackexchange.com/questions/184029/what-is-elastic-net-regularization-and-how-does-it-solve-the-drawbacks-of-ridge.

[18]“Regularization: Ridge, Lasso and Elastic Net,” *DataCamp Community*. Available: https://www.datacamp.com/community/tutorials/tutorial-ridge-lasso-elastic-net.

[19] "Select DOM Element using D3", *Tutorialsteacher.com*, 2019. [Online]. Available: https://www.tutorialsteacher.com/d3js/select-dom-element-using-d3js . [Accessed: 09- Jul- 2019].

[20] "D3 and Canvas in 3 steps", *Developer News*, 2019. [Online]. Available: https://www.freecodecamp.org/news/d3-and-canvas-in-3-steps-8505c8b27444/ . [Accessed: 15- Jul- 2019].

[21] "D3 in Depth", *D3indepth.com*, 2019. [Online]. Available: https://www.d3indepth.com/geographic/ . [Accessed: 21- Jul- 2019].

[22] "How to Perform Multiple Asynchronous Tasks with D3", *Jeffrey Morgan*, 2019. [Online]. Available: https://usabilityetc.com/2016/08/how-to-perform-multiple-asynchronous-tasks-with-d3/. [Accessed: 02- Jul- 2019].

[23] “Fetch", *Javascript.info*, 2019. [Online]. Available: https://javascript.info/fetch. [Accessed: 14- Jul- 2019].

[24] S. setInterval, "Scheduling: setTimeout and setInterval", *Javascript.info*, 2019. [Online]. Available: https://javascript.info/settimeout-setinterval. [Accessed: 17- Jul- 2019].

[25] "JSON.parse()", *W3schools.com*, 2019. [Online]. Available: https://www.w3schools.com/js/js_json_parse.asp. [Accessed: 20- Jul- 2019].

[26] "topojson/topojson", *GitHub*, 2019. [Online]. Available: https://github.com/topojson/topojson. [Accessed: 21- Jul- 2019].

[27] "D3 + simpleheat.js geographic heatmap", *Gist*, 2019. [Online]. Available: https://gist.github.com/patricksurry/803a131d4c34fde54b9fbb074341daa5. [Accessed: 02- Jul- 2019].

[28] "D3 in Depth", *D3indepth.com*, 2019. [Online]. Available: https://www.d3indepth.com/scales/. [Accessed: 19- Jul- 2019].

[29] "Multiple Line Climate Plot with Zoom", *Bl.ocks.org*, 2019. [Online]. Available: http://bl.ocks.org/natemiller/7dec148bb6aab897e561. [Accessed: 15- Jul- 2019].

[30] Y. Holtz, "Bubble chart template in d3.js", *D3-graph-gallery.com*, 2019. [Online]. Available: https://www.d3-graph-gallery.com/graph/bubble_template.html. [Accessed: 15- Jul- 2019].

[31] "d3 mouseover multi-line chart", *Bl.ocks.org*, 2019. [Online]. Available: https://bl.ocks.org/larsenmtl/e3b8b7c2ca4787f77d78f58d41c3da91. [Accessed: 22- Jul- 2019].

[32] Y. Holtz, "Brushing in d3.js", *D3-graph-gallery.com*, 2019. [Online]. Available: https://www.d3-graph-gallery.com/graph/interactivity_brush.html. [Accessed: 20- Jul- 2019].

[33] Y. Holtz, "Add tooltip to bubble chart in d3.js", *D3-graph-gallery.com*, 2019. [Online]. Available: https://www.d3-graph-gallery.com/graph/bubble_tooltip.html. [Accessed: 23- Jul- 2019].

[34] *D3js.org*, 2019. [Online]. Available: https://d3js.org/d3.v4.min.js. [Accessed: 01- Jul- 2019].

[35] *D3js.org*, 2019. [Online]. Available: https://d3js.org/d3-scale-chromatic.v1.min.js. [Accessed: 14- Jul- 2019].

[36] *D3js.org*, 2019. [Online]. Available: https://d3js.org/topojson.v1.min.js. [Accessed: 16- Jul- 2019].

[37] "mourner/simpleheat", *GitHub*, 2019. [Online]. Available: https://github.com/mourner/simpleheat/blob/gh-pages/simpleheat.js. [Accessed: 18- Jul- 2019].

[38] 2019. [Online]. Available: https://codepen.io/dodozhang21/full/KFkaf. [Accessed: 15- Jul- 2019].