PROJECT REPORT ON BIG DATA VISUALIZATION MADE EASY

Project Report

Presented to

Prof. Kevin Smith Department of

Computer Science San José State

University

In Partial Fulfillment

Of the Requirement for the Class

CS 235

By

Rashmeet Kaur Khanuja (012409982)

December, 2017

Abstract

"A picture is worth a thousand words" – especially when you are trying to find relationships and understand your data – the size of which may know no bounds, and the organization no structure. To create meaningful visuals of such data, there are a variety of design patterns to consider to make the visualization user-friendly. Data size and composition play an important role in this process. Our project aims to visualize humongous raw data which consists of US road traffic volume for the year 2015. This
~2gb dataset contains daily volumes of traffic, binned by the hour. The data used has the following fields: daily observation of traffic volume divided into 24 hourly bins, station_id, traffic flow direction, and type of road. Some fields we used as is, while the others we performed pre-processing and aggregation on to derive meaningful inferences. For visualization, we majorly used D3.js along with other libraries like datamaps, d3-svg-legend, d3-tooltip, and bootstrap (CDN). For data pre-processing, we used python. Using these technologies, we aim to create an interface for multiple user scenarios that is backed by strong analytics while providing a friendly-yet-informative dashboard to the users to identify trends to suit their needs.

Keywords : big data, road traffic, big data visualization, d3.js, datamaps, heatmaps, preattentive variables, bootstrap, road traffic data US

Table of Contents

1.0 Introduction

ROAD TRAFFIC is the flux or passage of motorized and non-motorized vehicles on the road. With the rapid development of transportation systems and integration of newer technologies with vehicles, traffic has become a vital part of human life and significantly influenced the quality of life. For instance, American drivers spend an average of more than 17,600 minutes behind the wheel each year, according to a new survey from the AAA Foundation for Traffic Safety [1]. With the growing population and large number of vehicles problems, like congestion, accidents and air pollution are increasing. A number of efforts to address these problems have been proven effective, including intelligent transportation systems (ITSs), public transportation systems, safety seat belts, and air-bags. Among these solutions, ITSs are deemed attractive because they enhance the efficiency and functionalities of transportation systems with advanced information technology. In particular, the role of data in ITS has become essential because of the size of the collected data. Traffic data refers to datasets generated and collected on moving vehicles and objects. Data visualization is an efficient means to represent distributions and structures of datasets and reveal hidden patterns in the data. Huge amount of numbers in big spreadsheets, no matter how meaningful, are not only tedious to go through but also get very time consuming and boring for the users. Presenting these numbers in a form that is visually appealing and informative as well provides a platform for the users to retrieve relevant results while cutting out the amount of time and efforts spent on manual, labor-intensive experimentation. Data visualization is not a new concept; a lot of work has been done in this field before. Visualization-based data discovery platforms allow business users to mash up disparate data sources to create custom analytical views with flexibility and ease of use that simply didn't exist with the raw data. "Watch 24 Hours of Traffic, Visualized as a Living Circulatory System" by Max Galka uses similar technologies namely D3.js and Three.js to visualize traffic on the highways as living circulatory system.[2] We also found the work done by Matteo Picozzi , Nervo Verdezoto , Matti Pouke , Jarkko Vatjus-Anttila and Aaron Quigley in their paper - *"Traffic Visualization - Applying Information Visualization Techniques to Enhance Traffic Planning"* relevant in terms of technologies they used namely D3.js along with highcharts and Google Maps API in order to represent geospatial data for exploratory visualizations. other related work in this field has also been done using real-time data that we shall elaborate upon in the future scope for our project section.

The dataset we chose here consists of road traffic volumes in US throughout the year 2015. Since our data file was in textual format, we used python to convert it into CSV (since it is compatible with D3.js). Furthermore, it converted the raw data into a more structured format to perform aggregations and other calculations for various derivations. Our objectives behind visualizing this data were -

- *Getting answers from the visualization of data such as, where are the heaviest traffic volumes? By time of day? Any interesting seasonal patterns to traffic volumes? Analyze traffic volumes during the holiday season? Traffic trends in rural and urban areas?*
- *Study of traffic routes between various cities within states*

- *Plot interactive maps of trends obtained from the data analysis, e.g. Busiest states during the year? Traffic in areas within states, e.g. - station ids? Direction of travel?*
- *Provide a dashboard suitable for multiple user scenarios so as to serve various user segments*
- *Learning new interactive data visualization technologies such as D3.js and datamaps.*

Our ultimate motivation for this project was to help users make data driven decisions through traffic visualization.

2.0 Functional Design

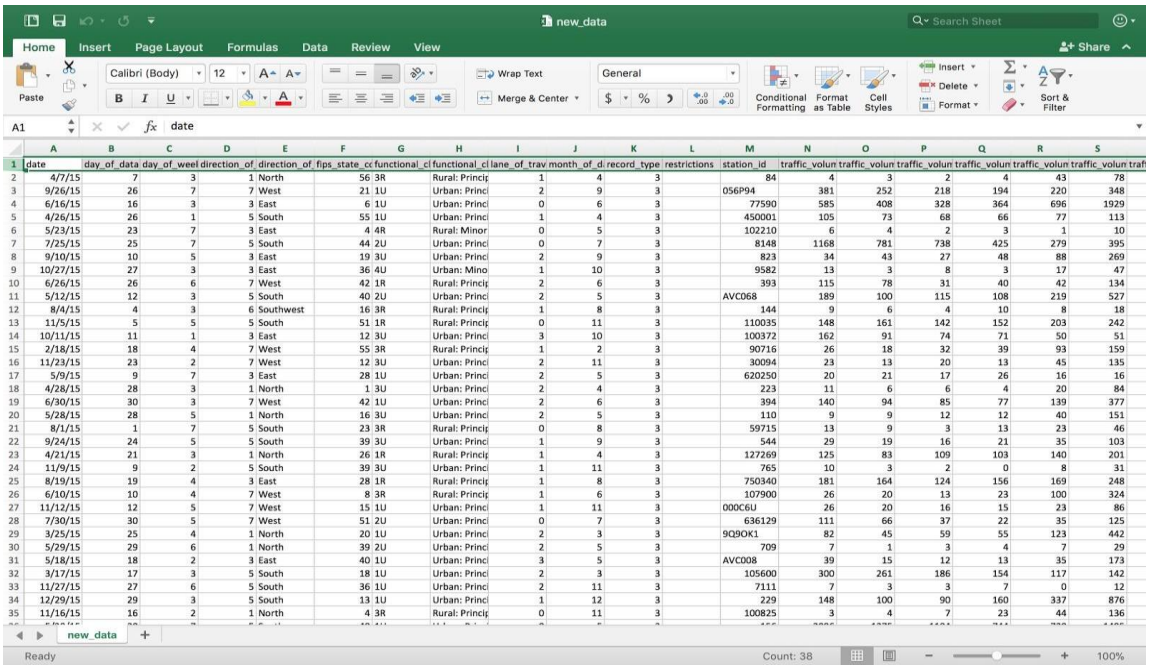
In this section, we shall discuss what our prototype does and how. We shall also cover the requirements and design approach.

Before we decided on the design approach, it was important to first understand the nature of data. We answered questions like how our data was organized, how the elements related to each other, how the user could operate the data (filtering, sorting, etc.), and how the details of our data could be focused upon. Answering these help us take steps towards deciding the functional design of our software. Our dataset was highly unstructured and needed cleaning, matching, and organizing; that called for a lot of data pre-processing. We chose python for the entire pre-processing step. First we used it to write scripts in order to change the data format from text (Fig. 1) to Comma Separated Values (Fig. 2). Furthermore, we added more understandable columns by mapping them with the available fields like state names corresponding to the fips code provided in the data set (Fig. 3), month names corresponding to numeric values (Fig. 4) day names like Sunday, Monday, and so on for values 1, 2, 3…given in our data set (Fig.

5) , and more . We further performed aggregation on data and added a column for the sum of traffic (Fig. 6). Also we used it to create a separate data file for the information of traffic volume during holidays(Fig. 7). Once our data organization was clear, the next step was to decide upon the design patterns to be used. In order to pick design patterns to represent our data, we also needed a clear picture of functional, non-functional, and user experience requirements.

```
Date,day_of_data,day_of_week,direction_of_travel,direction_of_travel_name,fips_state_code,functional_classification,functional_classification_name,lane_of_travel,month_of_data,record_type,restrictions,station_id,traffic_volume_counted_after_0000_to_0100,traffic_volume_counted_after_0100_to_0200,traffic_volume_counted_after_0200_to_0300,traffic_volume_counted_after_0300_to_0400,traffic_volume_counted_after_0400_to_0500,traffic_volume_counted_after_0500_to_0600,traffic_volume_counted_after_0600_to_0700,traffic_volume_counted_after_0700_to_0800,traffic_volume_counted_after_0800_to_0900,traffic_volume_counted_after_0900_to_1000,traffic_volume_counted_after_1000_to_1100,traffic_volume_counted_after_1100_to_1200,traffic_volume_counted_after_1200_to_1300,traffic_volume_counted_after_1300_to_1400,traffic_volume_counted_after_1400_to_1500,traffic_volume_counted_after_1500_to_1600,traffic_volume_counted_after_1600_to_1700,traffic_volume_counted_after_1700_to_1800,traffic_volume_counted_after_1800_to_1900,traffic_volume_counted_after_1900_to_2000,traffic_volume_counted_after_2000_to_2100,traffic_volume_counted_after_2100_to_2200,traffic_volume_counted_after_2200_to_2300,traffic_volume_counted_after_2300_to_2400,year_of_data
"2015-04-07","7","3","1","North","56","3R","Rural: Principal Arterial –
Other","1","04","3",,"000084","00084","00003","00002","00004","00043","00078","00116","00144","00132","00115","00150","00184","00169","00136","00129","00089","00122","00124","00110","00069","00073","00028","00012","00006","15"
"2015-09-26","26","7","7","West","21","1U","Urban: Principal Arterial –
Interstate","2","09","3",,"056P94","00381","00252","00218","00194","00220","00348","00453","00679","00826","00962","01158","01379","01376","01383","01453","01617","01669","01308","01068","00928","00885","00798","00650","00613","15"
"2015-06-16","16","3","3","East","06","1U","Urban: Principal Arterial –
Interstate","0","06","3",,"077590","00585","00408","00328","00364","00696","01929","04228","05634","05673","04636","03925","03827","04049","03954","04077","04244","04405","04609","04361","03272","02243","02050","01453","00892","15"
"2015-04-26","26","1","5","South","55","1U","Urban: Principal Arterial –
Interstate","1","04","3",,"450001","00105","00073","00068","00066","00077","00113","00254","00367","00487","00668","00870","00996","01003","01000","01043","01011","00959","00851","00708","00559","00457","00297","00207","00110","15"
"2015-05-23","23","7","3","East","04","4R","Rural: Minor
Arterial","0","05","3",,"102210","00006","00004","00002","00003","00001","00010","00017","00052","00064","00068","00082","00096","00099","00087","00087","00083","00061","00055","00035","00029","00021","00023","00009","00007","15"
"2015-07-25","25","7","5","South","44","2U","Urban: Principal Arterial – Other Freeways or
Expressways","0","07","3",,"008148","01168","00781","00738","00425","00279","00395","00715","01202","01486","01819","02193","02401","02691","02508","02700","02673","02746","02564","02257","02140","02060","02156","01873","01590","15"
"2015-09-10","10","5","3","East","19","3U","Urban: Principal Arterial –
Other","2","09","3",,"000823","00034","00043","00027","00048","00088","00269","00334","00476","00444","00355","00375","00434","00492","00499","00539","00542","00571","00559","00395","00351","00262","00217","00153","00061","15"
"2015-10-27","27","3","3","East","36","4U","Urban: Minor
Arterial","1","10","3",,"009582","00013","00003","00008","00003","00017","00047","00095","00158","00151","00151","00141","00184","00193","00219","00217","00238","00278","00250","00154","00130","00084","00056","00021","00014","15"
"2015-06-26","26","6","7","West","42","1R","Rural: Principal Arterial –
Interstate","2","06","3",,"000393","00115","00078","00031","00040","00042","00134","00268","00335","00392","00441","00600","00722","00652","00706","00751","00740","00784","00822","00703","00425","00192","00159","00150","00168","15"
"2015-05-12","12","3","5","South","40","2U","Urban: Principal Arterial – Other Freeways or
Expressways","2","05","3","AVC068","00109","00100","00115","00108","00219","00527","01212","01596","01347","01115","01114","01126","01178","01193","01364","01662","01673","01519","01148","00914","00808","00555","00468","00270","15"
"2015-08-04","04","3","6","Southwest","16","3R","Rural: Principal Arterial –
Other","1","08","3",,"000144","00009","00006","00004","00010","00008","00018","00047","00097","00116","00107","00128","00156","00143","00145","00150","00121","00140","00104","00096","00070","00060","00062","00031","00014","15"
"2015-11-05","05","5","5","South","51","1R","Rural: Principal Arterial –
Interstate","0","11","3",,"110035","00148","00161","00142","00152","00203","00242","00363","00562","00715","00713","00902","01005","00922","01031","01058","00999","01028","00881","00692","00584","00525","00431","00346","00282","15"
"2015-10-11","11","1","3","East","12","3U","Urban: Principal Arterial –
Other","3","10","3",,"100372","00162","00091","00074","00071","00050","00051","00110","00164","00250","00381","00532","00544","00620","00472","00484","00444","00419","00423","00433","00345","00311","00255","00308","00124","15"
"2015-02-18","18","4","7","West","55","3R","Rural: Principal Arterial –
Other","1","02","3",,"090716","00026","00018","00032","00039","00093","00159","00206","00251","00202","00246","00176","00228","00216","00226","00262","00260","00311","00314","00205","00148","00134","00091","00074","00037","15"
"2015-11-23","23","2","7","West","12","3U","Urban: Principal Arterial –
Other","2","11","3",,"030094","00023","00013","00020","00013","00045","00135","00429","00691","00713","00604","00581","00545","00540","00563","00507","00515","00483","00478","00324","00198","00097","00089","00079","00041","15"
"2015-05-09","09","7","3","East","28","1U","Urban: Principal Arterial –
Interstate","2","05","3",,"620258","00020","00021","00017","00026","00016","00016","00049","00094","00112","00104","00122","00108","00146","00166","00149","00192","00178","00167","00109","00106","00093","00100","00061","00033","15"
"2015-04-28","28","3","1","North","01","3U","Urban: Principal Arterial –
Other","2","04","3",,"000223","00011","00006","00006","00004","00020","00084","00397","00735","00466","00313","00335","00446","00469","00374","00374","00399","00395","00380","00276","00167","00131","00090","00034","00013","15"
"2015-06-30","30","3","7","West","42","1U","Urban: Principal Arterial –
Interstate","2","06","3",,"000394","00140","00094","00085","00077","00139","00377","00901","01486","01365","00925","00931","00869","00928","01051","01245","01291","01351","01521","00975","00640","00489","00338","00313","00204","15"
"2015-05-28","28","5","1","North","16","3U","Urban: Principal Arterial –
```


Fig. 1



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	date	day_of_data	day_of_week	direction_of	direction_of	fips_state_co	functional_cl	functional_cl	lane_of_trav	month_of_d	record_type	restrictions	station_id	traffic_volun	traffic_volun	traffic_volun	traffic_volun	traffic_volun	traffic_volun
2	4/7/15	7	3	1	North	56	3R	Rural: Princip	1	4	3		84	4	3	2	4	43	78
3	9/26/15	26	7	7	West	21	1U	Urban: Princi	2	9	3		056P94	381	252	218	194	220	348
4	6/16/15	16	3	3	East	6	1U	Urban: Princi	0	6	3		77590	585	408	328	364	696	1929
5	4/26/15	26	1	5	South	55	1U	Urban: Princi	1	4	3		450001	105	73	68	66	77	113
6	5/23/15	23	7	3	East	4	4R	Rural: Minor	0	5	3		102210	6	4	2	3	1	10
7	7/25/15	25	7	5	South	44	2U	Urban: Princi	0	7	3		8148	1168	781	738	425	279	395
8	9/10/15	10	5	3	East	19	3U	Urban: Princi	2	9	3		823	34	43	27	48	88	269
9	10/27/15	27	3	3	East	36	4U	Urban: Mino	1	10	3		9582	13	3	8	3	17	47
10	6/26/15	26	6	7	West	42	1R	Rural: Princip	2	6	3		393	115	78	31	40	42	134
11	5/12/15	12	3	5	South	40	2U	Urban: Princi	2	5	3		AVC068	189	100	115	108	219	527
12	8/4/15	4	3	6	Southwest	16	3R	Rural: Princip	1	8	3		144	9	6	4	10	8	18
13	11/5/15	5	5	5	South	51	1R	Rural: Princip	0	11	3		110035	148	161	142	152	203	242
14	10/11/15	11	1	3	East	12	3U	Urban: Princi	3	10	3		100372	162	91	74	71	50	51
15	2/18/15	18	4	7	West	55	3R	Rural: Princip	1	2	3		90716	26	18	32	39	93	159
16	11/23/15	23	2	7	West	12	3U	Urban: Princi	2	11	3		30094	23	13	20	13	45	135
17	5/9/15	9	7	3	East	28	1U	Urban: Princi	2	5	3		620250	20	21	17	26	16	16
18	4/28/15	28	3	1	North	1	3U	Urban: Princi	2	4	3		223	11	6	6	4	20	84
19	6/30/15	30	3	7	West	42	1U	Urban: Princi	2	6	3		394	140	94	85	77	139	377
20	5/28/15	28	5	1	North	16	3U	Urban: Princi	2	5	3		110	9	9	12	12	40	151
21	8/1/15	1	7	5	South	23	3R	Rural: Princip	0	8	3		59715	13	9	3	13	23	46
22	9/24/15	24	5	5	South	39	3U	Urban: Princi	1	9	3		544	29	19	16	21	35	103
23	4/21/15	21	3	1	North	26	1R	Rural: Princip	1	4	3		127269	125	83	109	103	140	201
24	11/9/15	9	2	5	South	39	3U	Urban: Princi	1	11	3		765	10	3	2	0	8	31
25	8/19/15	19	4	3	East	28	1R	Rural: Princip	1	8	3		750340	181	164	124	156	169	248
26	6/10/15	10	4	7	West	8	3R	Rural: Princip	1	6	3		107900	26	20	13	23	100	324
27	11/12/15	12	5	7	West	15	1U	Urban: Princi	1	11	3		000C6U	26	20	16	15	23	86
28	7/30/15	30	5	7	West	51	2U	Urban: Princi	0	7	3		636129	111	66	37	22	35	125
29	3/25/15	25	4	1	North	20	1U	Urban: Princi	2	3	3		9Q9OK1	82	45	59	55	123	442
30	5/29/15	29	6	1	North	39	2U	Urban: Princi	2	5	3		709	7	1	3	4	7	29
31	5/18/15	18	2	3	East	40	1U	Urban: Princi	3	5	3		AVC008	39	15	12	13	35	173
32	3/17/15	17	3	5	South	18	1U	Urban: Princi	2	3	3		105600	300	261	186	154	117	142
33	11/27/15	27	6	5	South	36	1U	Urban: Princi	2	11	3		7111	7	3	3	7	0	12
34	12/29/15	29	3	5	South	13	1U	Urban: Princi	1	12	3		229	148	100	90	160	337	876
35	11/16/15	16	2	1	North	4	3R	Rural: Princip	0	11	3		100825	3	4	7	23	44	136

Fig. 2

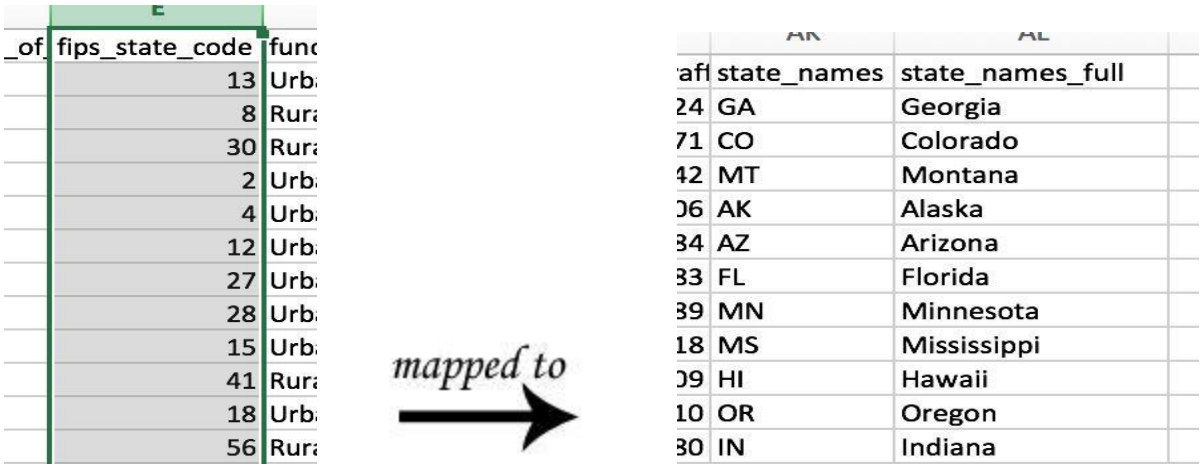


Fig. 3

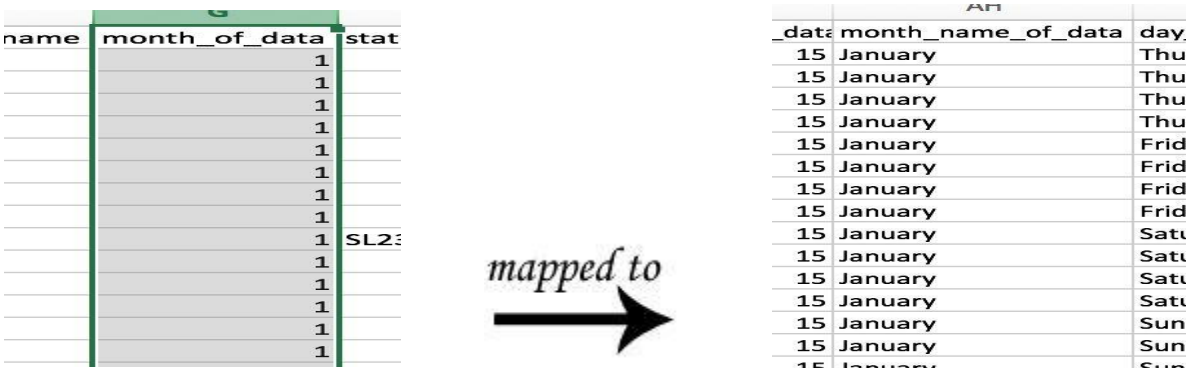


Fig. 4

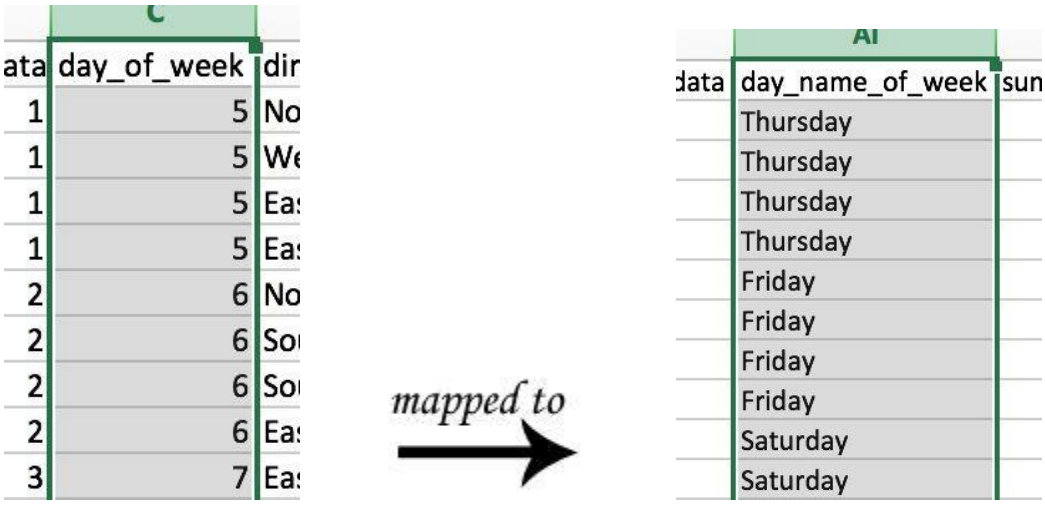


Fig. 5

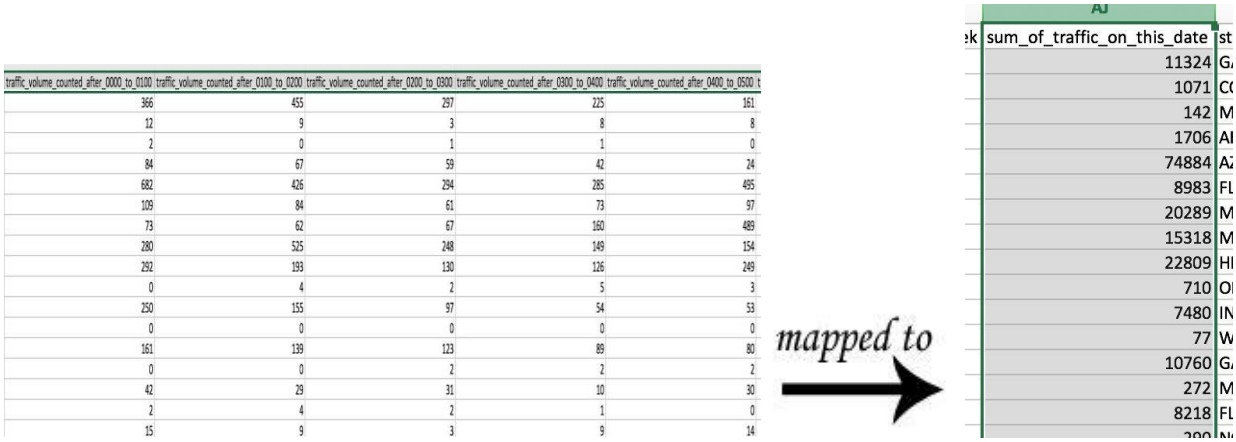


Fig. 6

Holiday Name	Traffic Volume	Traffic Volume on day
New Year's Day: Jan1	16560444	30-Dec
Martin Luther King Jr. Day: Jan 19	16560444	16-Jan
Valentine's Day: Feb 14	145000	13-Feb
Memorial Day: May 25	130000	22-May
Independence Day: July 4	100000	2-Jul
Labor Day: Sep 7	80000	4-Sep
Halloween: Oct 31	100000	29-Oct
Veterans Day: Nov 11	100000	12-Nov
Thanksgiving Day: Nov 26	100000	25-Nov
Christmas Eve: Dec 24	100000	23-Dec

Fig. 7

Requirements - In order to visualize past traffic data, decision makers need to be able to navigate through various fields of the data and be able to view different elements. They must be able to make sense of the visuals by looking at different views of the data such as traffic in states throughout the year, traffic distribution based on months, days, and hours, options to view traffic in rural and urban regions, and on holidays. One of the major functional requirements of our project was for the visualization to be interactive and flexible so as to provide the users with functionality to refine the views based on their requirements. The visualization needed to be informative enough to fulfill all these requirements, yet not too cluttered to prevent overwhelming the user.

Design Approach - Taking into account the objectives of traffic data analysis, the tasks of traffic data visualization can be classified as follows -

2.1 **Traffic Situations Monitoring** - In road traffic data, interesting trends may be hidden such as days with maximum traffic, routes with the busiest roads, stations that need more attention to avoid jams, pollution information, etc.

2.2 **Trend Search** - One of the most important goals of data visualization is to discover patterns. These patterns/trends reflect useful information such as "On Fridays, between 5pm-11pm, the route between San Francisco and San Jose shows maximum heat". This, if noticed on a repeated basis, can be termed as a trend and such trends can help various users in varied segments. It also helps discover the relationship between different parameters such as Day, Time, and Station Id with respect to Traffic Volume in the above example.

2.3 **Contingent exploration and forecasting** - Having data visualized to convey meaningful information, it can be used to study past trends and predict the future. For eg., taxi trips are found to be most frequent in the downtown on Friday nights after 10 pm. Such data can, for instance, be used by Uber drivers to plan their driving schedule for the upcoming weekend accordingly and make most profit.

2.4 **Route Selection Mechanism** - Human intelligence coupled with data vis insights can be used to plan and recommend routes in order to facilitate better analytics and improve traffic control efficiency .

We based our design approach on the above requirements and tasks that must be taken care of by an efficient visualization system. We brainstormed, surveyed, and analyzed various user-scenarios in order to incorporate elements that would benefit multiple user segments and provide a platform to visualize big data in an appealing way. We incorporated meta overview, interactive filtering, details on demand, macro and micro views, preattentive variables to focus user's subconscious attention, textual representation of relevant highlights, and layering and positioning to prevent overloading the user with information. We shall discuss about the use of various design patterns and their implementation in the upcoming sections.

3.0 System Design

Our system architecture provides the aforementioned interactive visualization dashboard. The high level architecture is composed of two components - the Data Processor and the View Generator (Fig. 8).

Input Data - Input data here refers to the raw data file that we obtained in text format. This dataset was unstructured and unorganized which is why it needed to be passed into the data processor in order to clean, map, and organize it. The original data consisted of 38 columns.

Data Processor - This component takes the raw input data and runs python scripts on that data to provide CSV files in proper structure. It also does the job of performing the necessary cleaning, rearranging, and sorting of data in order to make it usable according to the needed visualization input. **Online Processing** - This step consists of the various components where the front end work is done. The task of fetching the processed data and creating various design patterns based on it is accomplished here. It facilitates analysis and exploration of traffic variations flow using javascript libraries such as D3, datamaps, d3-svg-legend, and d3-tooltip along with html, css, and bootstrap to support the illustrations.

In this setup, some of the system requirements are as follows -

- Versions of languages used -

    - D3 V4
    - Datamaps 0.4.4 or higher
    - Python 2.7 or higher
- Memory requirement - minimum 8GB RAM
- IDE used for coding - Visual Studio Code 1.18, pyCharm 2017.3

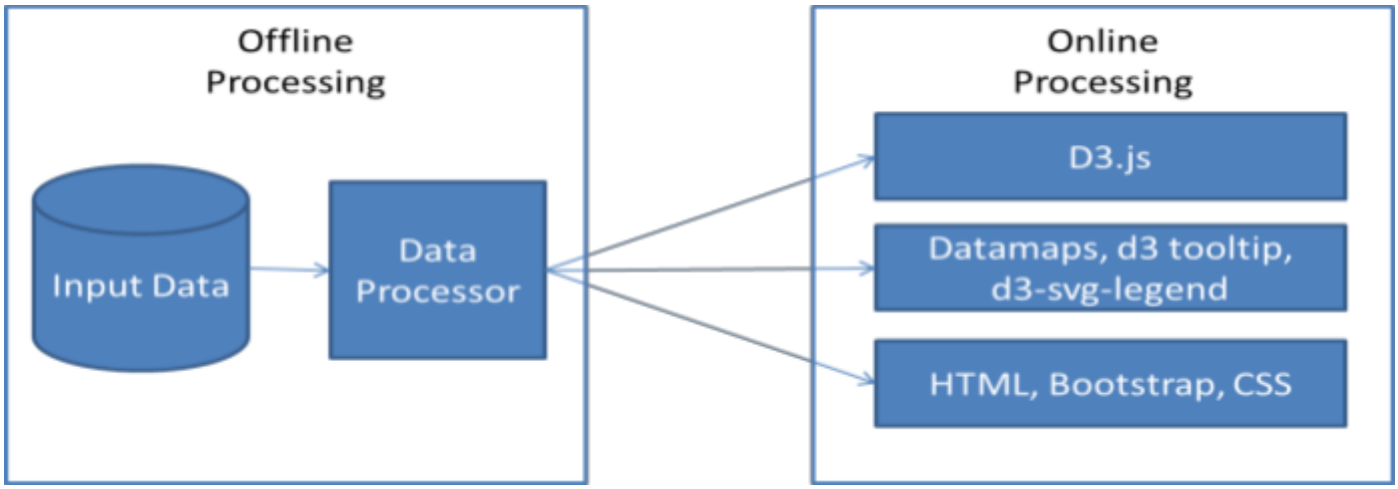


Fig. 8

4.0 Implementation

To show the different attributes of our data in a way that facilitates analysis and appeals the user, we decided to incorporate multiple **design patterns** to serve different needs. We needed to make sure that the information conveyed to the user reduces their cognitive load, hence we adopted the following approach.

4.1. For the primary view (Fig. 9), we chose **country heatmap** to display the distribution of traffic across different states in the US for the entire year. It was best suited for our data because displaying state-wise information along with additional details was required. The depth of colors on the map corresponds to the volume of traffic in that state; on mouse hover the state color changes to show it is in focus. Alongside the heatmap, we provided a drop down menu for data filtering. It enables the user to select the month for which they wish to see the traffic volumes. Data legend is used to represent data values based on different colors. To show points of interest, labels are used to display the traffic information of the state on which mouse is hovered. The preattentive variable used in heatmap is **color saturation.** There's a table below the dropdown menu to display useful highlights such as total traffic during the selected month, state with the highest traffic, and month with the highest traffic.

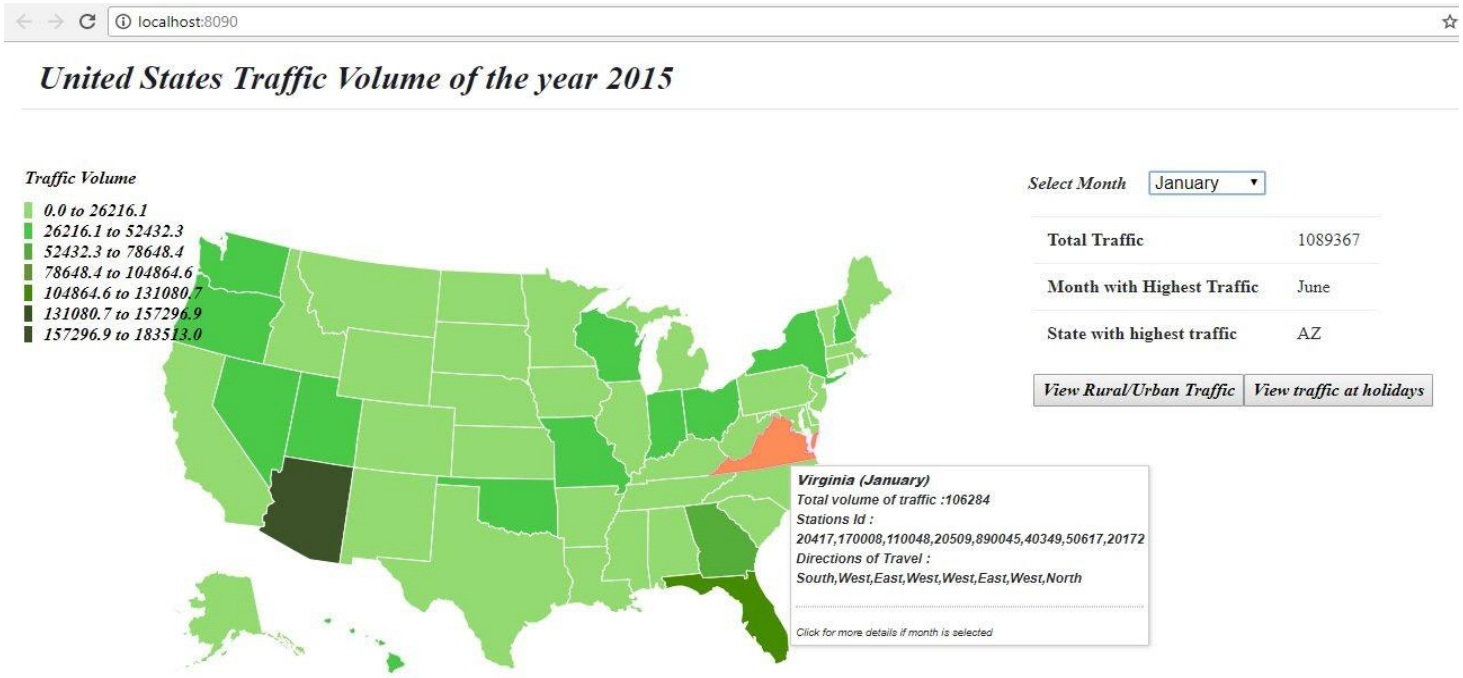


Fig. 9

4.2 When the user clicks on a particular state in the country heatmap, a modal opens that shows a **histogram** to represent the day-wise traffic of the selected state in the selected month (Fig. 10). To represent data values, we used numeric framing where X-axis shows the date and Y-axis the traffic volume. The height of the bars corresponds to the volume of traffic, thus the preattentive variable here being **size**. Upon mouse hover, the selected bar changes color and a tooltip appears that displays more information about the selected day's traffic count.

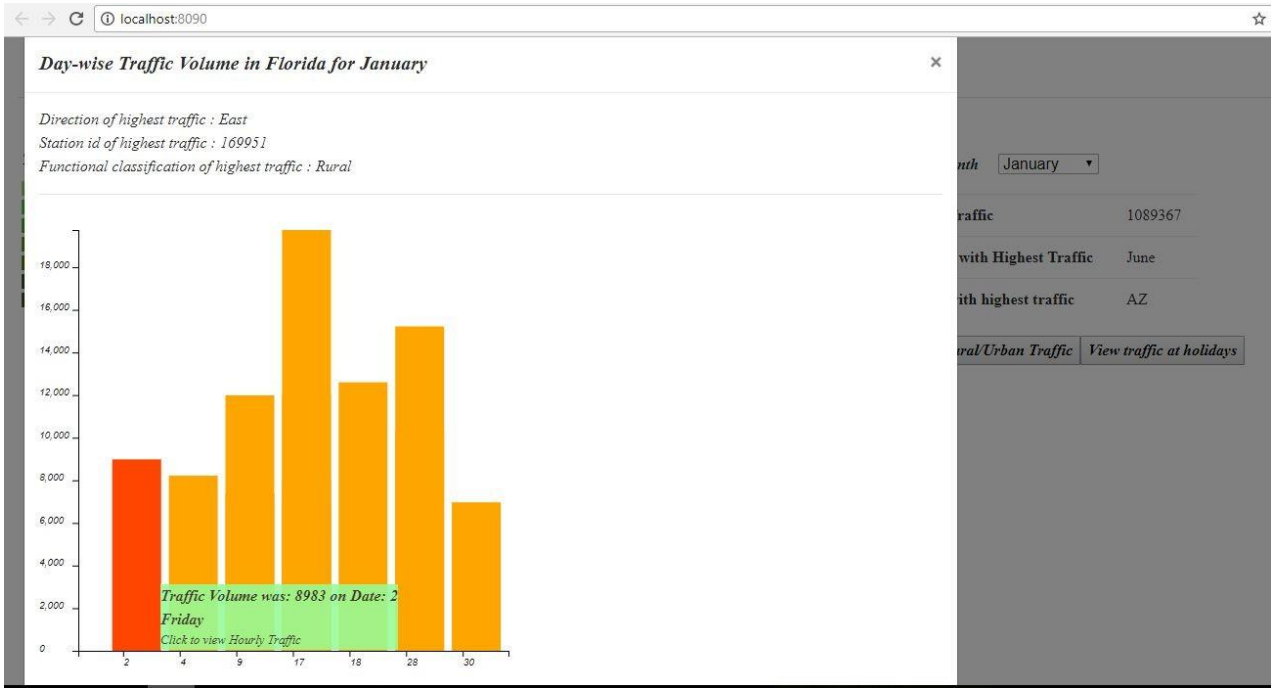


Fig. 10

4.3 If a user wishes to go deeper and view hourly traffic for a selected day in a particular state in the selected month, they can click on the day's bar and a **line graph** appears (Fig. 11). X-axis represents the

24 hours and Y-axis the traffic volume. Since the data points were more in number, line graph was the best fit to show peaks and valleys so as to make it easy for the user to just take a glance and infer peak traffic hours.

Fig. 11

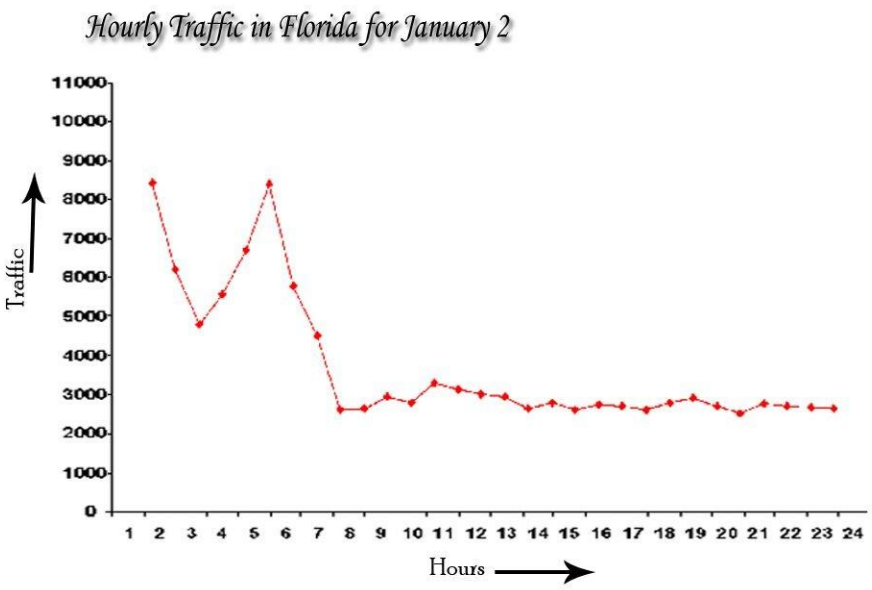

Hourly Traffic in Florida for January 2
Traffic
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Hours

Fig. 11

While the modal for the histogram or line graph is open, rest of the information such as the heatmap and other buttons are defocused in the background so as to let the user isolate relevant information.

4.4 To view traffic volumes in rural and urban, **pie chart** was our choice since it conveniently shows the comparison of traffic volumes between the two (Fig. 12). Both **color and size** were used as preattentive variables here.

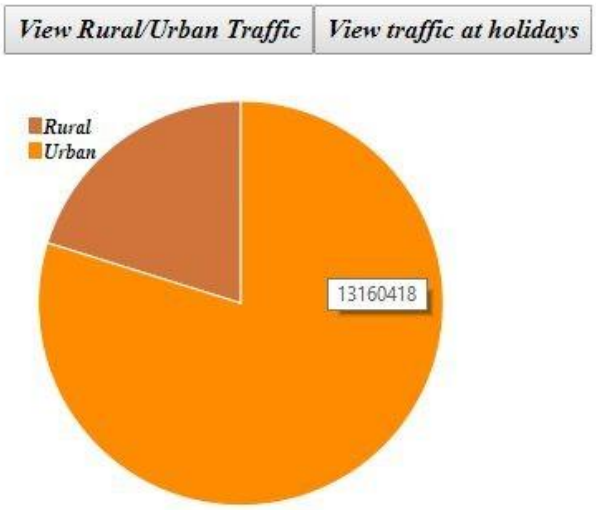


Fig. 12

4.5 We chose a **dynamic bubble chart** to represent the traffic data on holidays(Fig. 13). The size of bubbles corresponds to the volume of traffic, thus **size** being the preattentive variable. On mouse hover, titles appear that display the traffic volume and name of the holiday so as to let the users explore traffic trends during holidays.

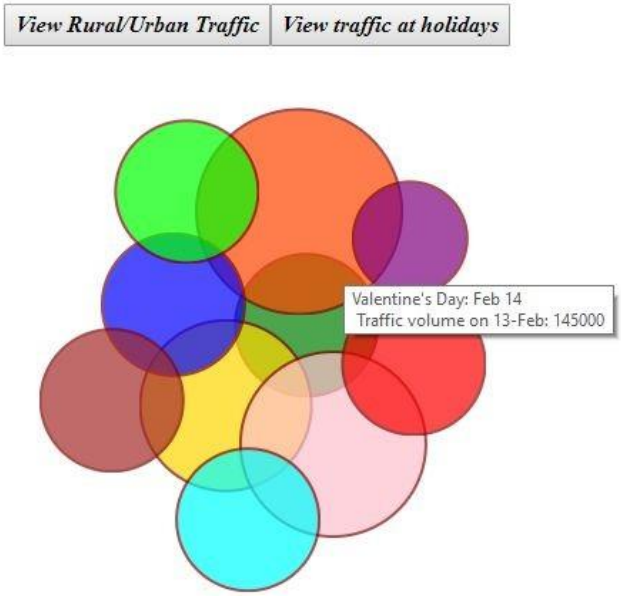


Fig. 13

4.6 Finally we used **dynamic queries** for filtering and sorting at various levels to provide atomic visualization options to the users.

All these views enable various levels of details to provide adequate information and their layering and positioning helps in decluttering so as not to make it difficult for the user by providing too much information at once.

This functional design is aimed to serve multiple user segments. Some examples of **user scenarios** that could benefit by using our dashboard are -

- common citizens - who want to view traffic details
- data analysts - to study road traffic trends
- cab companies - to plan their strategies
- traffic control department - to understand the traffic distribution and likelihood of congestion, and to manage assigning force accordingly

5.0 System Testing

To validate that the requirements we had set initially were being met and that our user-centered design framework goal was achieved, we performed continuous **usability testing** on our system. We approached the design process in a way to make it user friendly and keep the visible options at a single time limited while maintaining the delivery of relevant information.

Rather than testing the system in entirety upon completion, we adopted **unit testing** approach wherein we kept testing the working of our dashboard at various stages to check how various elements perform. The first phase was to test it ourselves in order to check each functionality. While testing the system for proper working of the country heatmap, we figured that it would be a good idea if we incorporated similar heatmaps for each state on mouse click. Since we had already implemented a bar chart on mouse click that appeared in a modal, and it was getting really difficult to open two modals at the same time, we prepared a wireframe mock-up of the state heatmaps. The idea was to show not only the traffic volume heat in different counties and stations within a state, but also the routes connecting these places and differentiating between the busy and light ones (Fig. 14).

Fig. 14

Upon testing more functionality, we discovered that even though we had performed data pre-processing at the onset itself, we still needed to use python for more data processing during development stages as well in order to add more columns, drop the columns that we found were irrelevant during the course of development, and to perform further aggregations.

Further we performed **integration testing** to ensure that the combination of various elements worked fine. In order to do this we selected a month from the drop down menu to check if the heatmap changes colors accordingly, the values in legend update on the basis of that month's data, the label values updated, and the overview table showed the correct highlights. In this process we discovered that the label values show too much information when no month is selected. While we tried to remove this feature for the yearly data (when none option is selected in the drop down menu), it was getting removed from the rest of the options as well (when a month was selected). Therefore we chose to let the labels be the same for all options.

Moreover, we performed **system testing** by running our dashboard on different operating systems namely Windows 10 and Mac OS. It turned out that the dashboard worked consistently on all of these. The load-time and refresh rate varied slightly depending upon the configuration of the systems.

One area where our system did not perform as expected was during **regression testing**. When the data size was increased above 0.5 GB, javascript would run out of heap memory and the dashboard wouldn't load. We tried fixing this issue but our attempts didn't seem to work. Finally we worked with a small data size for now since this project was mainly focused on the User Interface Design.

Finally we performed **acceptance testing** with the help of our colleagues. When the dashboard was presented to users, we got various types of feedback that helped us make some useful changes in order to make the GUI more user friendly. Some of the changes implemented at this stage were as follows -

- Placement of buttons to provide better clarity
- Removing the heatmap scrollbar - initially we had added a scrollbar to implement some functionality but that appeared to the users as if the drop down menu to select months and the heatmap were different sections, whereas they were supposed to be one section as a whole. To avoid this confusion, we removed the scrollbar and provided a complete page scroll functionality instead
- Levels of complexity - since there is a lot of information that can be inferred from our dataset, it was important to display views in a way that it doesn't overwhelm the user. Keeping this in mind, we used the "details on demand" approach wherein limited elements are visible at once and the user can click or hover to go into the deeper layers of functionality.
- Finally, we made some minor changes like pleasing fonts, vibrant colors, combination of colors to provide appeal yet avoid distraction, and so on.

6.0 Conclusion and Future Work

Big Data presents numerous opportunities to draw inferences and study trends but is also cumbersome to process and analyze in it's raw form. Thus here we presented a working dashboard that visualizes US road traffic data for the year 2015. Our visualization addresses problems arising from large scale, multi-modal, unstructured data and provides comprehensible visuals to the users that are appealing to the eyes and make the job of analysis and forecasting easy and fun. We adopted various design patterns to show different aspects of data along with the corresponding pre-attentive variables to focus the user's attention on the highlights. The levels of complexity in our dashboard have been taken care of by displaying views on different layers. We performed various types and levels of testing including feedback from users from different segments and incorporated changes accordingly so as to make the GUI more user-friendly.

In future work, we will address augmentations that can be done to our current implementation - (i) Using large size dataset by storing it in a database like MongoDB or other and then writing web services to fetch the data and visualize it using dynamic queries. (ii) Using real-time data instead of static data so as to let the decision-makers view current road traffic conditions and plan their strategies accordingly. Kafka can be used for this purpose. (iii) Adding another vertical by creating an app for our dashboard so as to make this analytics platform mobile and provide the decision-makers on the go services (iv) Currently, the dashboard refers to road traffic only, further air-traffic and other data can also be used for similar visualizations that will help in horizontal growth.

References

1. AAA NewsRoom. (2017). *Americans Spend an Average of 17,600 Minutes Driving Each Year | AAA NewsRoom*. [online] Available at: http://newsroom.aaa.com/2016/09/americans-spend-average-17600-minutes-driving-year/ [Accessed 3 Dec. 2017].
2. Metrocosm. (2017). *Watch 24 Hours of Traffic, Visualized as a Living Circulatory System - Metrocosm*. [online] Available at: http://metrocosm.com/map-us-traffic/ [Accessed 3 Dec. 2017].
3. Chen, W., Guo, F. and Wang, F. (2015). A Survey of Traffic Data Visualization. [online] Available at: http://www.cad.zju.edu.cn/home/vagblog/VAG_Work/IEEEITS2015/TrafficVASurvey.pdf [Accessed 2 Dec. 2017
4. Picozzi, M., Verdezoto, N. and Pouke, M. (2013). Traffic Visualization Applying Information Visualization Techniques to Enhance Traffic Planning. [online] Available at: https://pure.au.dk/ws/files/70285046/TrafficVisualizationIVAPP2013.pdf [Accessed 2 Dec. 2017]
5. Data link: https://kaggle.com
6. d3 library: http://d3js.org/d3.v4.js
7. Other libraries:https://maxcdn.bootstrapcdn.com/bootstrap/4.0.0-alpha.6/css/bootstrap.min.css
8. Other libraries: https://code.jquery.com/jquery-3.1.1.slim.min.js
9. Other libraries:https://cdnjs.cloudflare.com/ajax/libs/tether/1.4.0/js/tether.min.js
10. Other libraries:https://maxcdn.bootstrapcdn.com/bootstrap/4.0.0-alpha.6/js/bootstrap.min.js
11. Datamaps library:http://datamaps.github.io/
12. Python IDE: https://www.jetbrains.com/pycharm/
13. Visual Studio Code software: https://code.visualstudio.com/download
14. Code referrred: https://github.com/sakethsaxena/leading-causes-of-death-US-app