Theme Park Accident Dynamic Web Dashboard: Write-up

SMU Data Analytics Bootcamp | Project 3

Group 1: April Gao, Will Hughes, Vanessa Martinez, and Omar Zaher

I. Introduction

The subject for our group was inspired by the recent news that Frisco, Texas was decided as the location for a new Universal Studio family friendly theme park. Excited, we began to do research on theme parks in general and found that the information out there was only regarding accidents on different types of thrill rides. Most of our data was contained to an 8 year time span ranging between 2000-2008. We realized quickly that these parks aren't as safe as we would like to believe and the “Theme Park Capital of the World” is located in a state that is not required to report instances of accidents on typical theme park rides. We created a web based dynamic dashboard that displays visual graphs of the answers to our research questions we came up with using the data we found.

- What type of theme or amusement park (business type) has more reported incidents of injury?
- How many accidents were reported over time?
- What type of rides (device types) cause the most amount of injuries?
- Where are the parks with reported accidents located?

II. Data

Our data came from a database dedicated to rides, courtesy of Saferparks. Saferparks is a non-profit public service organization founded in 2000 to help prevent amusement ride accidents

through research, information sharing, and effective public safety policy. The organization is no longer actively engaged in research or advocacy, and provides the data they collected to the public for free use. Our original data-set came in csv format with 14,885 rows. We used Pandas and Jupyter Notebook to clean the data and narrow our dataframe to a sample size of 1000 rows, which uses a mask to range the incidents reported between 2000-2008. These parks are not required to report their names, so we also created a new column labeled “address” so we could create a map with the incident locations. Majority of the accidents in our dataset included the State where each accident was reported in, but a lot of the Cities were missing values. Subsequently the latitude and longitude coordinates were not provided. We used a Geocoding API from OpenWeatherMap.org to request the latitude and longitude for the given Cities and States. There were around 600 locations with both City and State provided in our dataset.

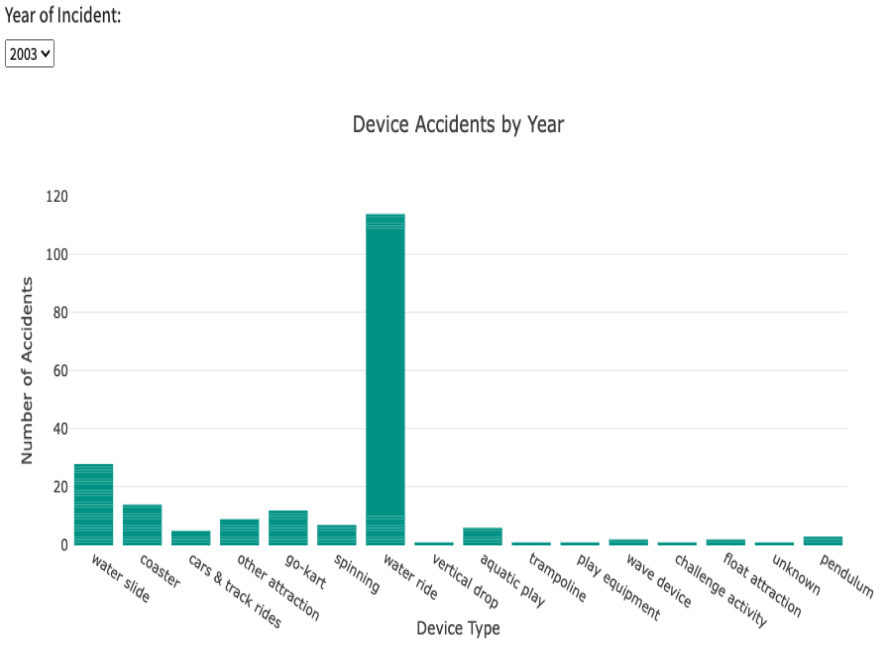
The main columns we use in our sample dataframe are “acc_id, acc_date, address, year, injury_desc, acc_desc, num_injured, device_type, bus_type”. The column titles stand for accident id, accident date, address, year, injury description, accident description, number injured, device type, and business type. We named our new cleaner data “sample_df” and saved it as a .json file to use in VS Code to create our dynamic dashboard.

III. Visualizations & Design

We built our dynamic website dashboard using HTML, JavaScript, and CSS. Our websites consists of 5 pages; the Main page displaying our 3 graphs, a Map page, Data page, About Us, and Works Cited page The main libraries we imported into our HTML page for our data visualization charts were Plotly, d3, Tableau, Leaflet, jquery, and Popper. It is dynamic in

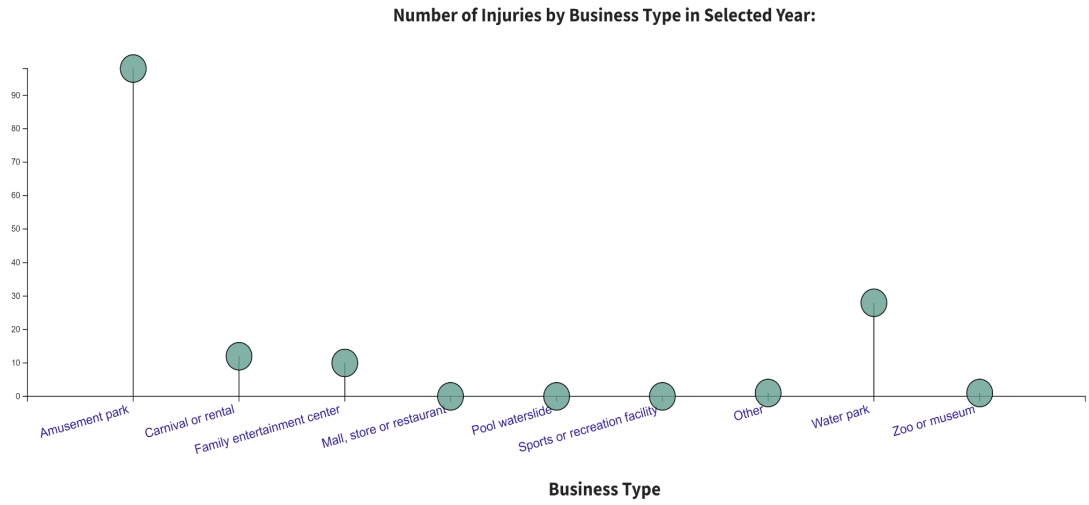
the sense that when you filter using the button above our graphs, all the graphs will adjust accordingly to the filter buttons.

Using Plotly, we were able to graph a bar chart that looked at injury count by device type and filtered by year. We found that 2003 had the highest number of reported injuries and water rides had the highest rate of injury, closely followed by coasters.

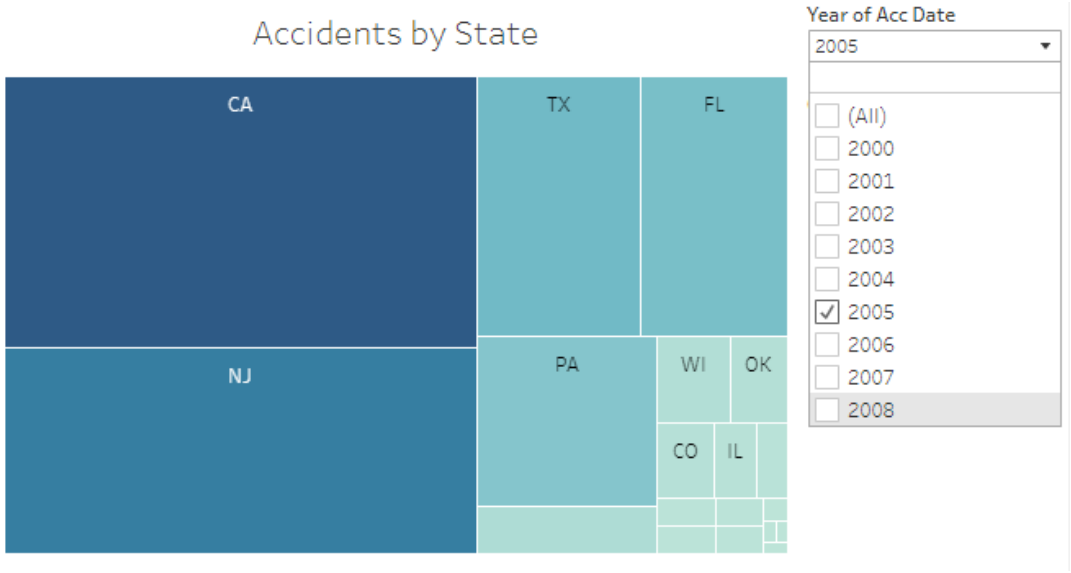


We also created a d3 lollipop chart analyzing the number of injuries grouped by business types and year. We first used Jupyter Notebook to group the data by year and business type, to calculate the number of injuries in each year. We dropped the unnecessary data that was not needed for the lollipop chart . We then moved “sum of injuries by year” from rows to columns, to have the 8 business types on rows and the columns showing the total number of injuries by year. We then saved the data to a CSV file for the d3 chart. We used the CSV file to create the lollipop chart. From the lollipop chart below, Amusement parks show the most injuries overall compared to the 8 other business types from 2000-2008. Waterparks ranked second in majority of the years, while zoos and museums show that they are the safest from looking at the chart. Of

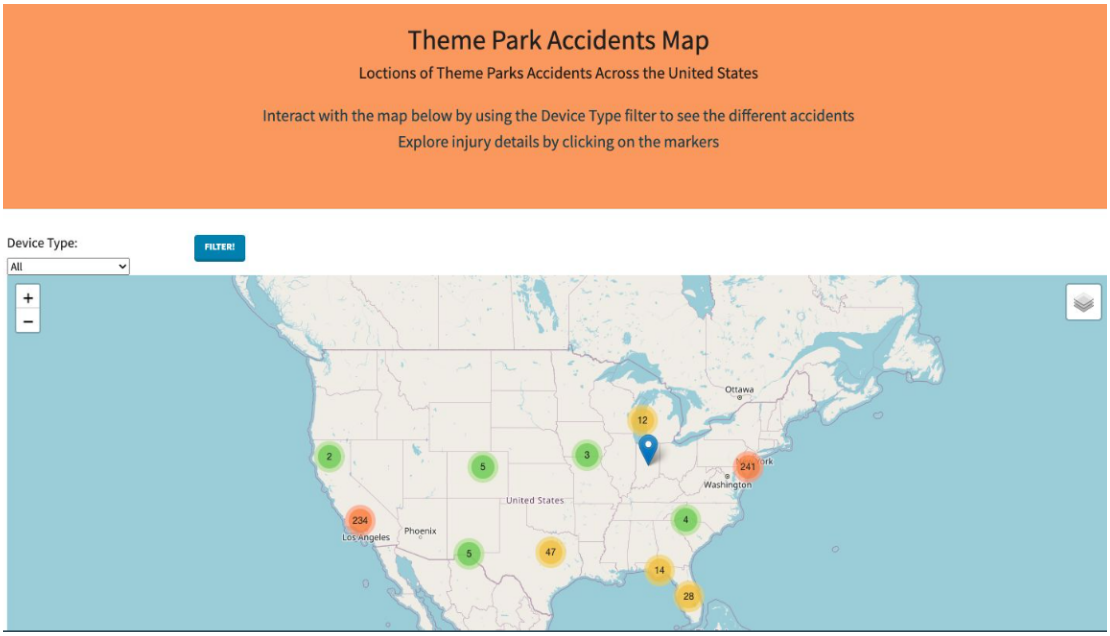
course this is just a fraction of the total theme parks accidents. We understand the data limitations and for the data collected, zoos and museums are the safest theme parks.

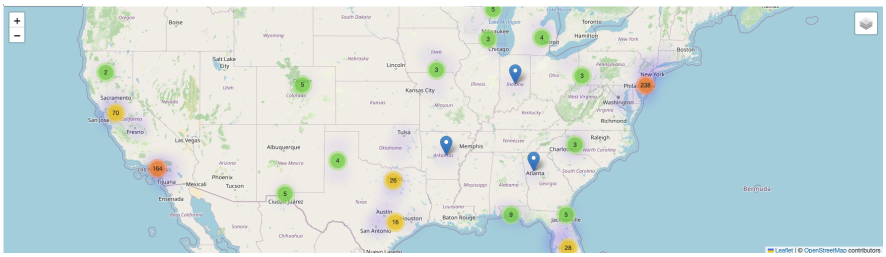


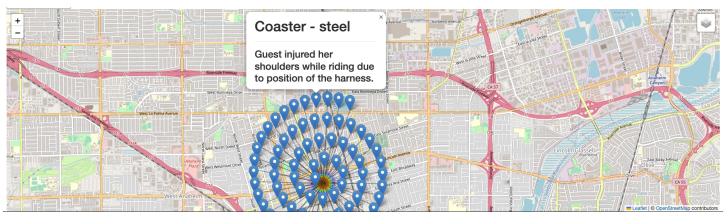
In addition to Plotly and D3, we also added a Tableau Tree Map to display the share of injuries broken down by state. This chart highlights how California and New Jersey make up more than half of all the injuries overall. It also allows the user to see just how few of the accidents are reported in Florida. One of our expected outcomes was seeing Florida at the front of the pack as it is home to so many Disney and Universal Theme Parks, but instead it is usually found somewhere in the middle. Florida isn't required to report all of their roller-coaster accidents and therefore appears to have fewer accidents than we anticipated.

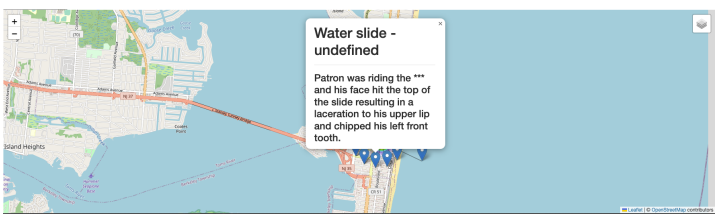


Our website's Map page was created using a Leaflet map with a heat filter and marker clusters. The interactive map shows accidents by device type and the injury description associated with the accident. The Map page initially loads displaying all of the locations from our data with a zoomed out perspective of the United States. The geographic coordinates for the center of the U.S. allow the map to be centered when the page is first shown. The view can be displayed as "Street" or "Topography" and the markers and heatmap can be selected on or off. The filter button allows the map to be interactive by letting the user select a device type, to then see where and if that type of device had reported accidents and descriptions. Some of the descriptions were vague and some had asterisks to replace the name of the device. The majority of reported accidents occurred in California or New Jersey.





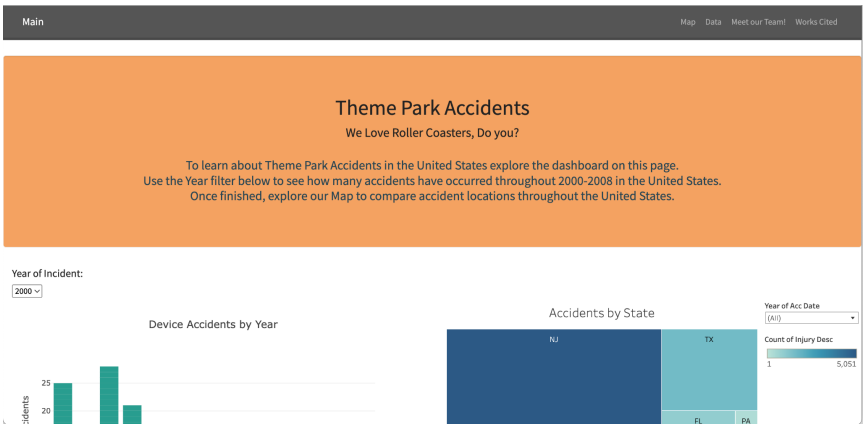


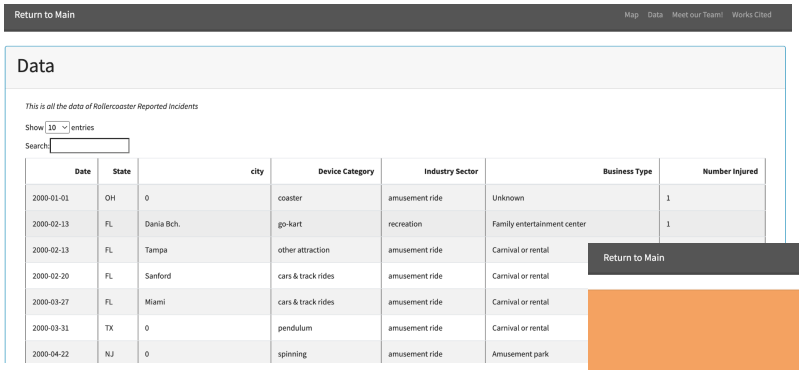


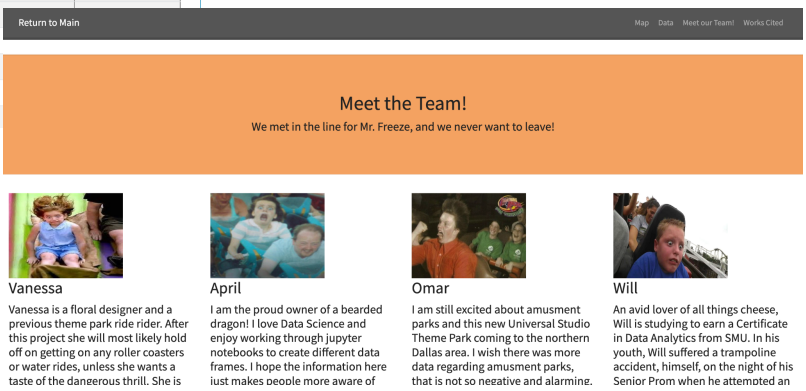
Our website's design and color palette initially was going to be a "retro" type theme. We realized that finding colors and designs to represent theme or amusement parks, in general, is not as easy as one would think. If the parks were more specified, such as Universal or Disney, the colors could potentially be blue and green, to represent a globe that accompanies Universal, or red, black and yellow to represent Mickey Mouse, associated with Disney. The parks in our data represent a wide variety of business types and since our information was focused on accidents we didn't want the design to be morose or gloomy. Icon images of roller coasters and ferris wheels were used via Canva. The colors used were orange, peacock green, and navy gray, with a gray navigation bar.

Color Palette

Website Images







IV. Conclusions

In conclusion, some of the most popular theme parks in the world include Disneyland, Disney World, Universal Studios, and Six Flags. These parks attract millions of visitors each year, who come to experience the excitement of their favorite rides and attractions, as well as to enjoy the atmosphere and ambiance of the park itself. The number of accident reports for theme parks depend on many variables unrelated to rider safety including popularity of the ride type, regulatory inclusions/exclusions, local government record retention and public disclosure policies, and individual corporate record keeping policies. States with stronger government oversight tend to log more accidents. States that carefully monitor a broad range of safety incidents, have efficient data management systems, and provide transparency to the public will, by definition, produce a higher number of public accident reports. This is an indicator of more attention to safety, not less. There are a number of challenges faced by theme parks, including safety concerns, competition from other forms of entertainment, and the need to constantly innovate and improve their offerings to stay relevant. Despite all of this, we all should do our due diligence regarding our “safety” and bring awareness to the lack of accountability and reporting at these parks, when things may not be as safe and entertaining as they appear.

V. Future Work & Limitations

Future work we have in mind starts with finding a better color scheme that is more universal with amusement parks. The limitation with this is there are many different colors associated with them and we were not able to decide on the best color set to represent our data. We will organize our bar chart and lollipop chart in a descending order; d3 is a very complex library that requires some time to navigate and operate. To make our site more visually pleasing,

we will fit all the graphs to fit above the fold. Our JumboTron takes up a pretty decent amount of space so we will decrease that in size and change the size of our containers to fit the screen properly. The heat map filter displays a blue color rather than red so we need to update that so it engages our user and gets the correct message across. Limitations of our data include the fact that “Saferparks' accident data set may not reliably predict nation-wide or industry-wide patterns. The relative frequencies of certain types of accidents, on certain types of equipment, may not accurately reflect the aggregate safety records of all amusement devices in the United States. For example: State laws require that go-kart accidents be reported in Florida, but not in California. Thrill ride accidents at major theme parks must be reported in California, but not in Florida. Therefore, records from the Florida Dept. of Agriculture will tend to show a higher percentage of go-kart accidents and a lower percentage of roller coaster accidents, than is accurate for that state.” Another major limitation is that the parks don't have to report their DBA’s and as of today there are no more associations or any type of authority keeping up with the records of these accidents.

VI. Works Cited

- **Kaggle - Roller Coaster Accidents**: https://www.kaggle.com/datasets/stevenlasch/roller-coaster-accidents?select=safer-parks-accident-dataset.csv
- **Data – Rides Database**: https://ridesdatabase.org/saferparks/data/
- **Saferparks Accident Data**: https://ridesdatabase.org/wp-content/uploads/2020/02/Saferparks-data-description.pdf
- **Palette - Coolors**: https://coolors.co/palette/264653-2a9d8f-e9c46a-f4a261-e76f51
- **Canva**: https://www.canva.com/