**Sonika Prakash Soni** identified a data set based on the accidents in states of the USA from 2016 – 2021. [US\_Accidents\_Dec21\_updated.xlsx](https://1drv.ms/x/s!AmfFXwvkIOnRjiOjFHncrwjPmYUi?e=Acl5k0)

The proposed research question was –***How do we predict the accidents to reduce the number of accidents in the United States?***

the **BIG IDEA** worksheet- GPD 3

Identify a project you are working on where you need to communicate in a data-driven way

WHO IS YOUR **AUDIENCE**?

1. List the primary groups or individuals to whom you’ll be communicating.

Insurance agents and adjusters, vehicle owners and operators, state, regional, and local government agencies, emergency response agencies, auto bodies, mechanic shops, and towing companies.



(3) What does your audience care about?

Whether the Pacific is considered a hot spot for accidents, consumers will be better able to determine the cost of ownership of their vehicle regarding insurance and deductible costs they would potentially face, as well as what type of vehicle they choose to own from a safety standpoint.

1. If you had to narrow that to a *single person*,

whom would that be?

A government agency

(4) What action does your audience need to take?

They can look at the data and visualizations that will be presented to them, and make informed decisions on the cost of insurance, accident likelihood, what type of safety features they consider, and for government agencies, what they can do to make traffic flows and their streets and highways safer. Insurance adjusters will be able to determine pricing more accurately for their consumers and what coverages would best suit their markets.

WHAT IS **AT STAKE**?

What are the *benefits* of your audience acting in the way that you want them to?

They will better understand accident trends in their area and make better decisions regarding insurance pricing and coverage options, vehicle purchases, and the cost of owning those vehicles and ultimately making owning and operating vehicles safer in any given locality.

What are the *risks* if they do not?

Insurance costs will not reflect the area and its trends, coverage options will not be optimal, consumers will purchase unnecessary or inadequate coverage as well as vehicles that are too expensive to own in their region, governments will develop inadequate traffic options and street designs accelerating accident numbers, which will create more costs for the agency they represent which ultimately impacts their citizens in forms of taxes. The bottom line, without understanding the data and trends within it, all involved lose money.

FORM YOUR **BIG IDEA**

It should:

1. articulate your point of view,
2. convey what’s at stake, and
3. be a complete (and single!) sentence.

Our big idea is to better inform all parties concerned with this data, as described in the “who is your audience section” so they may better target issues related to vehicle accidents, provide cost-effective solutions to the problems, provide better insurance costs and coverages, and effectively select vehicles that are safe, effective, and inexpensive to own and maintain.

1. *Create a Tableau Workbook (also referred to as a project). Connect to your Excel file. Prep the selected dataset and conduct some initial data prepping and some initial data exploration using Tableau. -* DONE

Chart

Description automatically generated

1. *Define the grain and the measurement event of your dataset(s).*

*Grain-* it is an instance of the accident location and other variables like Stop/Exit/Weather conditions during that accident.

The grain for this dataset involves the exact longitude and latitude of when the accident happened, as well as the city, State, and Country location. Along with this, the grain includes weather characteristics and the start and end time of the actual event.

*Measurement of Event-* The number of accidents that took place in a certain county with the severity of the accident. The measurement event for our dataset is going to be the number of car accidents that happen in the Pacific States of the US, including California, Nevada, Washington, and Oregon

1. *Update a data dictionary using Excel for each dataset. Please refer to the* [Sample Data Dictionary (Excel)](https://webcampus.unr.edu/courses/88234/files/9776430/download?wrap=1) *and review the article* [Data dictionary: A how-to and best practices Links to an external site](https://medium.com/@leapingllamas/data-dictionary-a-how-to-and-best-practices-a09a685dcd61) *by Carl Anderson*. – Please refer to the excel sheet attached to the Assignment.
2. *Complete the* [Big Idea Worksheet (PDF)](https://webcampus.unr.edu/courses/88234/files/9776453/download?wrap=1)  / [Big Idea Worksheet](https://webcampus.unr.edu/courses/88234/files/10133355?wrap=1) [Actions](https://webcampus.unr.edu/courses/88234/assignments/1070262?module_item_id=2307003)  *(Word* *Doc) and storyboard (include it as an appendix). Refer to* [Module 5 Learning Resources](https://webcampus.unr.edu/courses/88234/pages/module-5-learning-resources) *for additional information*. -[*Please refer to page 1*](#_top)*.*
3. *State the research questions you will be examining to achieve your Big Idea. \*(Note: You can review the* [Activity: Generating a Research Question](https://webcampus.unr.edu/courses/88234/modules/351513) *again to refresh your memory regarding the elements of a good research question.)*

* How do weather conditions affect the severity and frequency of car crashes near the Pacific area?
* How does the distance travel between the start and end of the crash affect the severity of crashes in the Pacific area?
* How does the time of day affect the likeness of a crash in the Pacific area?

1. *List the kinds of Tableau visuals you will be creating to investigate your research questions and tell your story.*

* Map
* Line
* Bar
* Packed Tables
* Dual Lines
* Dual Combination
* Pie-Chart

**GPD 4**

**GPD 4**

**Part A.**

Each of the reports we generated in Tableau allows for a better representation of the data set we have selected. Many of the reports we have created a deal with various weather and visibility patterns across the western U.S. Others demonstrate the total number of accidents in given counties and/or cities and expand on those facts with information pertaining to expanded accident details. This approach is very beneficial as it allows for more precise discussion and investigation. These reports work well in conjunction with other reports focused on weather patterns as they can be used to show correlations between accidents and weather. Other reports generated are based on the number of accidents in cities and states and detail the overall severity of any given accident. These are useful to our project because of their ability to visually represent areas where more severe accidents occur, representing a greater trend, and can be used in conjunction with expanded accident detail and weather reports to best represent specific occurrences and outliers within the data. Finally, reports have been generated that show the number of accidents by county separated by daylight and nighttime hours. This report is useful in determining a correlation between time of day, visibility, and accident numbers. Of these different reports we have created, five stood out from the rest, and we have determined that they best represent and assist in answering our research question and best tell the story of our big idea.

**Part B.**

The first Viz we have selected is due to its ability to show average visibility and the number of vehicle accidents in the same areas. This is useful because we wish to show a possible correlation between visibility and accidents to insurance adjusters and government officials.

The second viz we have selected shows the total number of vehicle accidents separated by county. This viz allows for the cursor to have hovered over any county in our data area and displays the county name and the total number of crashes. This is useful to our research question because it aids in determining if any given area is a hotspot relative to other areas represented within our data set.

The third viz selected details by county the number of accidents and their time-of-day occurrence. In essence, the viz allows for the cursor to hover over any given county within a state and information on the county name, the number of accidents, and the day or night hours is displayed. Below this map, the same can be accomplished with the only difference being the number of accident occurrences in the night hours. This is very useful and helps to determine specific details about perceived accident hotspots and allows for greater data dissection to determine other correlations.

The fourth Viz we have selected provides a filtration option to select a specific city or city to display their accident volume. This is useful when determining perceived hotspots, and after a specific city has been selected more data is made available on the total number of accidents such as a description of the accident, the accidents location within the selected city, and the specific weather pattern reported on the day and in the location of any given accident. This benefits our big idea as adjustors and government officials could show details of accidents in greater detail.

The fifth and final Viz we have selected to complete our efforts in our Big Idea filters and displays average temperature by county across our data area. This is very useful because temperatures can affect road conditions and when used in conjunction and comparison to specific weather patterns and precipitation levels over a given period, shows an accurate representation of how weather impacts accident levels. It can be assumed that our data displays that the northwest is traditionally colder, and experiences more precipitation than California or the southwest, but it is yet to be determined if there are more accident occurrences relative to population metrics in the northwest at this time. The combination of the five Vis’s we have selected best provides the answers to the questions and Big Idea we have sought to find.

**GPD 5**

**GPD 5 Story Board**

This Viz (Morgan 1) Represents the average number of vehicle crashes and the average visibility filtered by each state where an individual accident occurs. This viz is useful for showing how visibility impacts driving conditions and the rate at which accidents occur. Accidents occur more in areas with low visibility, both in the northwest and in the L. An area, both areas are known for their limited visibility. L.A. due to smog and the northwest for fog and rain.

This viz (Sonika 2) represents the total number of accidents that occur in each state, filtered by the average level of precipitation. As evidenced by the first slide, California deals with a high number of accidents and a higher average precipitation level relative to other states in the region, close to the level of precipitation seen in the northwest, another hotspot for accidents.

This viz (Lukas 3) demonstrates the total number of accidents that occur within each county of the data set and filters them by the number of occurrences between day and nighttime driving conditions. There is a lower level of incidents that occur during the night than day, which seems to correlate with a larger number of drivers on the road, all of which can be affected by the different weather and precipitation patterns at any given time within an area. This further supports our claim of the correlation between driving conditions and accident levels.

This viz (Jared 1) shows the details of all accidents that take place in any given city within the data set. Each accident can be investigated to determine the overall location, time, area, and weather conditions. By applying the parameters within this viz it can be noted that there is a correlation between precipitation and weather and accident rates increasing despite their location on the map and an even higher occurrence in areas such as southern California and the Pacific Northwest.

This vis (Justen 2) demonstrates the average temperature of each county within the data set and fails to show any correlation between temperature and accident rates, though it could be argued that inclement weather can result in more accidents and this weather often comes at lower temperatures, though this viz is useful simply to inform the audience that areas may deal with colder weather and lower accident levels or vice versa, such as southern California or the northwest region. It is important to note that poor weather does not automatically result in more accidents, nor colder weather individually.

**GPD 5 Part 3**

**Dashboard 1:** The first dashboard that we created was made with the idea of the primary causes of accidents in mind. These include things such as the severity of the crashes based on the state, how the humidity of the crashes was affected by weather, the effects of visibility on the number of crashes, as well as the number of accidents that happen overall per county. Also, as far as the storyboard goes, it shows how This is pertinent to our Big Idea in the sense that in our effort to create a system that allows rideshare services such as Uber and Lyft to evaluate the dangers of the roads for their drivers, it also gives insight into ways that they can avoid certain elements of liability that we discussed in our Big Idea, which can help save their company on one of the major expenses that they have, which would be the legal implications of being an employer for effectively individual contractors on the road with their cars. Evaluating this data would not only help them lower the chances of risk to drivers but lower the potential prices of insurance due to the knowledge of safe driving practices.

**Dashboard 2:** This dashboard is a representation of the effects of temperature and distance on the average severity of the car crash. This is shown in the first viz (Justen2) through the average temperature of the county with each reported car crash and can be cross-referenced with the second viz (Morgan2) to show the average severity of the crash as well as the average distance that is traveled during the crash and how temperature plays a factor in these two measures. This can also be seen in our storyboard, in which we can see examples in the viz such as Morgan1 and Justen2 again. These visualizations help outline the importance of understanding the risk involved in driving in certain temperatures. As we can see from the first viz, most of the areas with the hottest average temperature are located close to the bottom of the map, something that should be expected, but what we notice when we cross-reference it to the other viz is that a lot of these areas often have lower average distance traveled when compared to places that have a lower temperature, which often has higher severity and distance traveled. These two factors together show that driving in a place with a lower temperature can increase the risk of there being a more severe car crash, which is often the case especially when there is a long distance traveled. As it pertains to our Big Idea, once again this dashboard’s focus on the potential risks of driving in lower and higher temperatures can allow ridesharing companies to evaluate where they need to be more focused on reducing their liability and chances for a crash, so focusing heavily on driver safety in areas in the Pacific Northwest would benefit them greatly. Also, this focus could lead to fewer crashes and less insurance35 liability for them because they could encourage their drivers to take more time and precaution when driving to limit the chances of there being a crash, especially a severe one.