Queries

Query 1: Top 10 States with the Most Accidents

```
SELECT State, count(*) AS Number_of_Accidents FROM `group-project-data-
225.bidataset.bi_dataset_fact` JOIN `group-project-data-
225.bidataset.bi_dataset_dim_location` USING (Location_ID)
GROUP BY State
ORDER BY Number_of_Accidents DESC
LIMIT 10
```

Query 2: Number of Accidents based on Severity

```
SELECT Severity, count(*) AS Number_of_Accidents FROM `group-project-data-
225.bidataset.bi_dataset_fact` GROUP BY Severity ORDER BY Number_of_Accidents DESC
```

Query 3: Average Length of Road Affected Based on Severity

```
SELECT Severity, AVG(Distance_mi) AS Average_Distance_Affected FROM `group-project-
data-225.bidataset.bi_dataset_fact` GROUP BY Severity ORDER BY
Average_Distance_Affected DESC
```

Query 4: Most Common Weather Conditions at the Time of Accident

```
SELECT Weather_Condition, count(*) AS Number_of_Accidents FROM `group-project-data-
225.bidataset.bi_dataset_fact` WHERE Weather_Condition IS NOT NULL GROUP BY
Weather_Condition ORDER BY Number_of_Accidents DESC
```

Query 5: Number of Accidents at Day/Night Time

```
SELECT Sunrise_Sunset, count(*) AS Number_of_Accidents FROM `group-project-data-
225.bidataset.bi_dataset_fact` WHERE Sunrise_Sunset IS NOT NULL GROUP BY
Sunrise_Sunset ORDER BY Number_of_Accidents DESC
```

Query 6: Number of Accidents Per Day of Week

```
SELECT FORMAT_DATETIME("%A", Start_Time) AS Day_of_Week, count(*) AS
Number_of_Accidents FROM `group-project-data-225.bidataset.bi_dataset_fact` GROUP BY
Day_Of_Week ORDER BY Number_of_Accidents DESC
```

Query 7:Number of Accidents by Hour of Day

```
SELECT FORMAT_DATETIME("%H", Start_Time) AS Hour_of_Day, count(*) AS
Number_of_Accidents FROM `group-project-data-225.bidataset.bi_dataset_fact` GROUP BY
Hour_of_Day ORDER BY Number_of_Accidents DESC
```

Query 8: Number of Accidents by Month

```
SELECT FORMAT_DATETIME("%B", Start_Time) AS Month, count(*) AS Number_of_Accidents
FROM `group-project-data-225.bidataset.bi_dataset_fact` GROUP BY Month ORDER BY
Number_of_Accidents DESC
```

Query 9: Number of Accidents by Year

```
SELECT FORMAT_DATETIME("%Y", Start_Time) AS Year, count(*) AS Number_of_Accidents FROM
`group-project-data-225.bidataset.bi_dataset_fact` GROUP BY Year ORDER BY
Number_of_Accidents DESC
```

Query 10: Number of Accidents by Timezone

```
SELECT Timezone, count(*) AS Number_of_Accidents FROM `group-project-data-
225.bidataset.bi_dataset_fact` JOIN `group-project-data-
225.bidataset.bi_dataset_dim_location` USING (Location_ID)
WHERE Timezone IS NOT NULL
GROUP BY Timezone
ORDER BY Number_of_Accidents DESC
```

Query 11: Average Visibility Based on Severity

```
SELECT Severity, AVG(Visibility_mi) AS Average_Visibility FROM `group-project-data-225.bidataset.bi_dataset_fact` GROUP BY Severity ORDER BY Average_Visibility DESC
```

Query 12: Top 10 Streets with the Most Accidents

```
SELECT Street, count(*) AS Number_of_Accidents FROM `group-project-data-225.bidataset.bi_dataset_fact` JOIN `group-project-data-225.bidataset.bi_dataset_dim_location` USING (Location_ID)
GROUP BY Street
ORDER BY Number_of_Accidents DESC
LIMIT 10
```

Query 13: Top 10 Cities with the Most Accidents

```
SELECT City, count(*) AS Number_of_Accidents FROM `group-project-data-225.bidataset.bi_dataset_fact` JOIN `group-project-data-225.bidataset.bi_dataset_dim_location` USING (Location_ID)
GROUP BY City
ORDER BY Number_of_Accidents DESC
LIMIT 10
```

Query 14: The Relative Side of the Street (Right/Left) in Address Field at the Time of Accident.

```
SELECT Side, count(*) AS Number_of_Accidents FROM `group-project-data-225.bidataset.bi_dataset_fact` WHERE Side IS NOT NULL GROUP BY Side ORDER BY Number_of_Accidents DESC
```

Query 15: The average temperature affected based on severity

```
SELECT Severity, AVG(Temperature_F) AS Average_Temperature_Affected FROM `group-project-data-225.bidataset.bi_dataset_fact` GROUP BY Severity ORDER BY Average_Temperature_Affected DESC
```

Query 16: Most frequent visibility at the time of accident

```
SELECT Visibility_mi, count(*) AS Number_of_Accidents FROM `group-project-data-225.bidataset.bi_dataset_fact` WHERE Visibility_mi IS NOT NULL GROUP BY Visibility_mi ORDER BY Number_of_Accidents DESC
```

Insights obtained from the queries:

- The state with the most number of accidents is California, followed by Florida and Texas
- The vast majority of accidents have a “Severity” level of 2 .
- Accidents with a severity level of 4 affected the greatest extent of the road with an average distance of 1.45 miles. Severity levels 2 and 3 affected a similar distance.
- The most common weather condition at the time of the accident was “Fair”. (1,107,194). The next most common conditions were “Mostly Cloudy” (363959) and “Cloudy” (348,767). This suggests that adverse weather conditions are not a major contributing factor to car accidents.
- The majority of accidents occur during the day (~63.7%)
- Friday had the most number of accidents. Saturday and Sunday saw the fewest number of accidents by a significant margin compared to every weekday.
- The highest number of accidents occurred between 5:00 PM and 6:00 PM, followed closely by between the time of 4:00 and 5:00 PM, and 3:00 PM to 4:00 PM.
- December had the most accidents (473943), followed by November (360696) and October (299131).
- 2021 had the largest number of accidents by a significant amount compared to other years at 1,511,745, followed by 2020 (625,864), and 2019 (258,615).
- Most accidents occurred in the Eastern time zone, followed by Pacific, Central, and Mountain.
- For each accident severity level, the average visibility was greater than 9 miles (with a maximum of 10). This suggests that visibility is not a major factor in the severity of a car accident.
- I-95 N had the most number of accidents, followed by I-5 N and I-95 S.
- The city with the most number of accidents is Miami with (4%)
- The relative side of the accident street is mostly Right(2353309) over Left(492032)
- The average temperature of lowest severity 1 is 71 Fahrenheit, average temperature for highest severity 5 is 58 Fahrenheit. Generally the severity goes up as temperature gets lower.
- The most frequent visibility during accidents is 10 mile with a total of 2230276 records.