BUSINESS REQUIREMENTS VIZ AND SQL VALIDATION

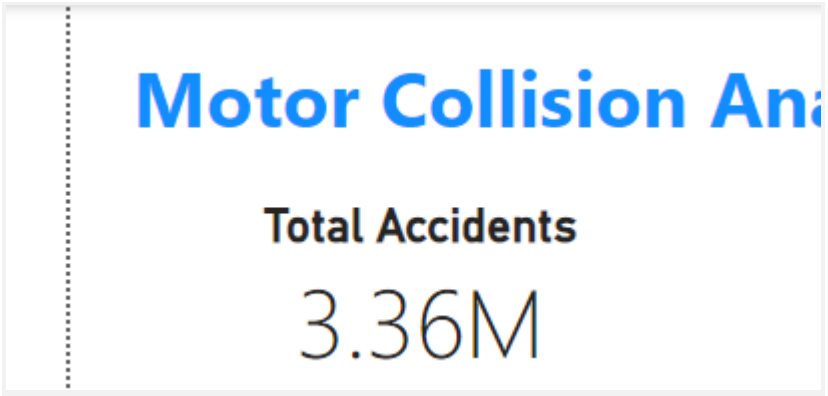
1. How many accidents occured in NYC, AUSTIN and CHICAGO?

→ SQL validation:

```sql
---Total number of accidents Overall:
SELECT COUNT(DISTINCT(COLLISION_ID)) FROM FACT_COLLISION; --3.36M
```

Results | Chart

	COUNT(DISTINCT(COLLISION_ID))
1	3356984

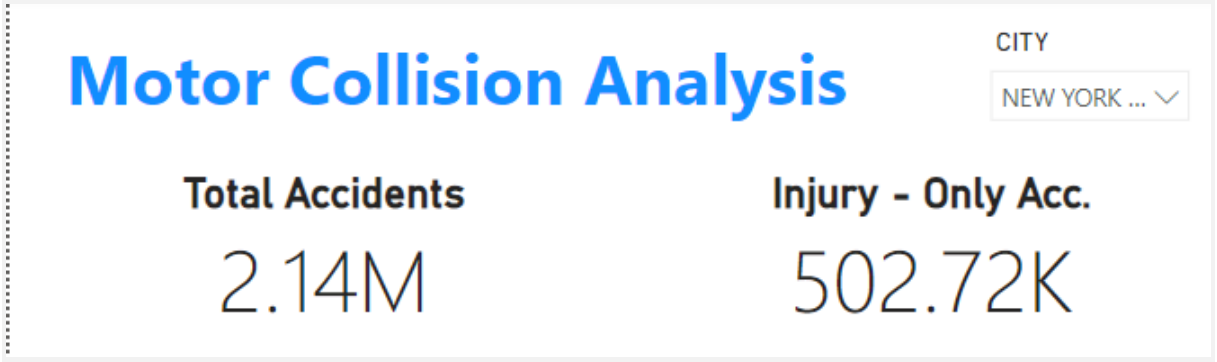


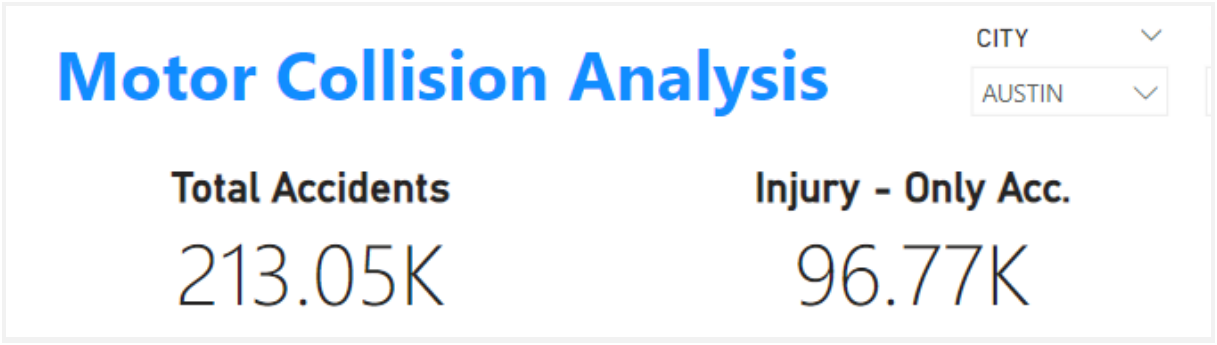
By City:

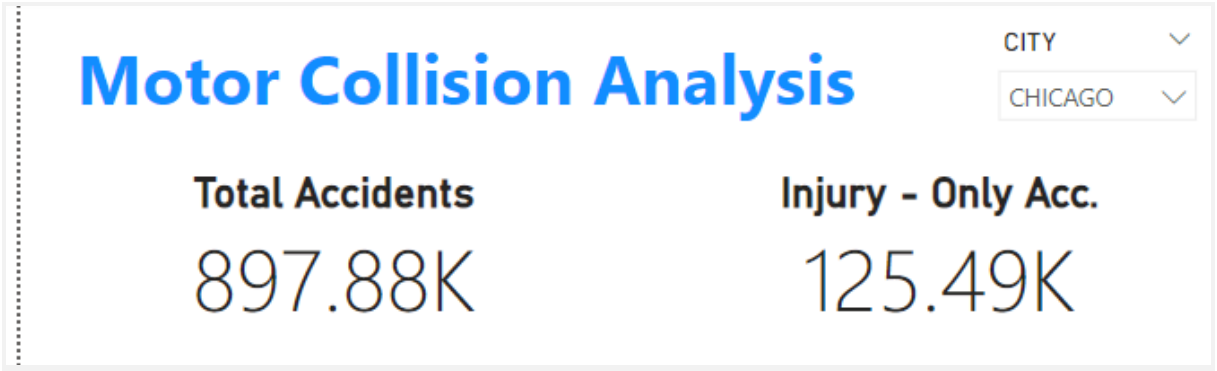
```sql
SELECT
    address.CITY,
    COUNT(DISTINCT fact.COLLISION_ID) AS Total_Accidents
FROM
    FACT_COLLISION AS fact
JOIN
    DIM_ADDRESS AS address
ON
    fact.ADDRESS_SK = address.ADDRESS_SK
WHERE
    address.CITY IN ('NEW YORK CITY', 'AUSTIN', 'CHICAGO', 'MONTGOMERY')
GROUP BY
    address.CITY;
```

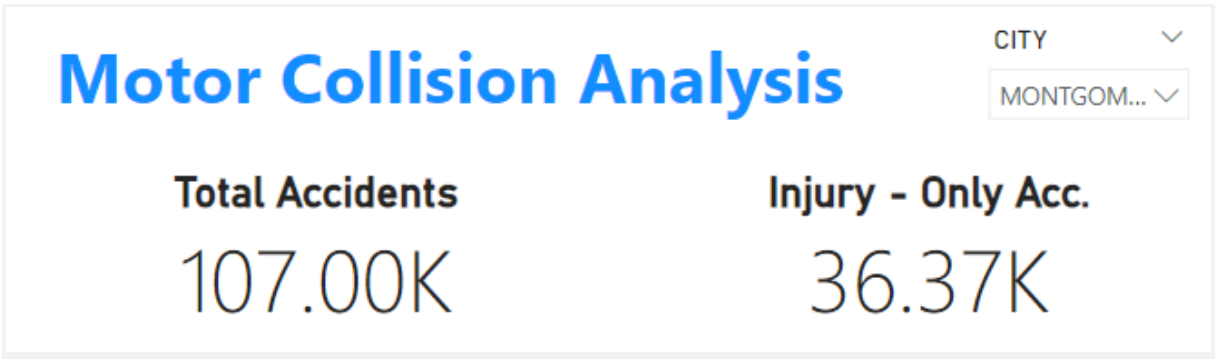
Results | Chart

	CITY	TOTAL_ACCIDENTS
1	NEW YORK CITY	2139048
2	AUSTIN	213053
3	CHICAGO	897880
4	MONTGOMERY	107003









2. Which areas in the cities has the greatest number of accidents? Top 3 areas in each city.

SQL:

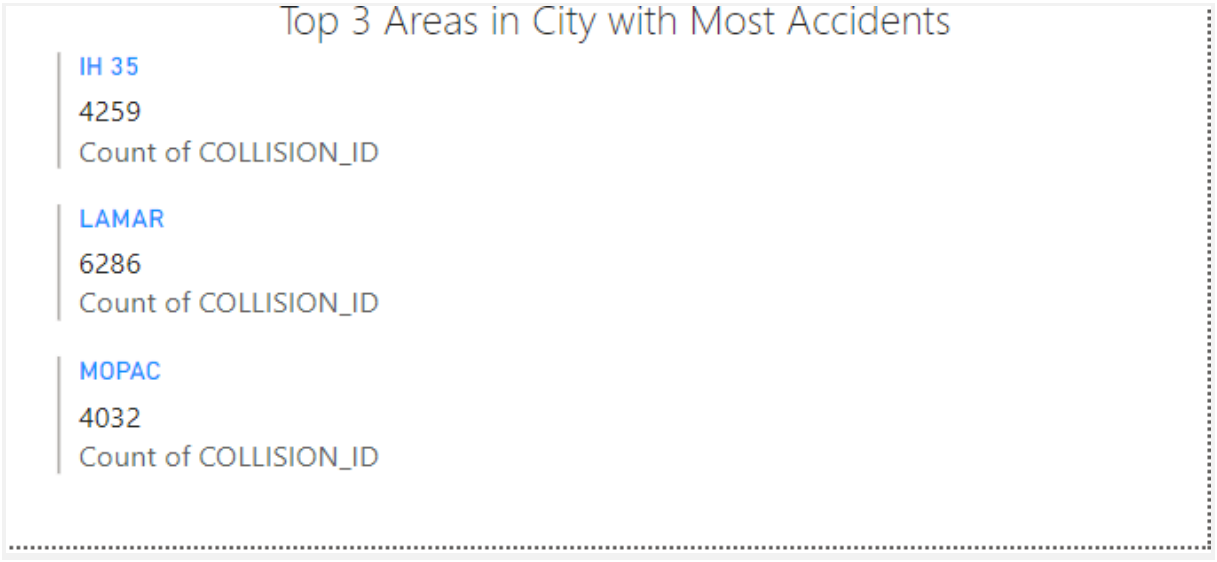
```
-- Top 3 areas in each city having most no.of accidents
WITH CityAreaAccidentCount AS (
    SELECT
        da.City,
        da.Street_Name AS Area,
        COUNT(DISTINCT fc.Collision_ID) AS Accident_Count
    FROM
        FACT_COLLISION fc
    JOIN
        DIM_ADDRESS da ON fc.Address_SK = da.Address_SK
    WHERE
        da.City NOT IN ('NA', '') AND da.Street_Name NOT IN ('NA', '')
```

Results | Chart

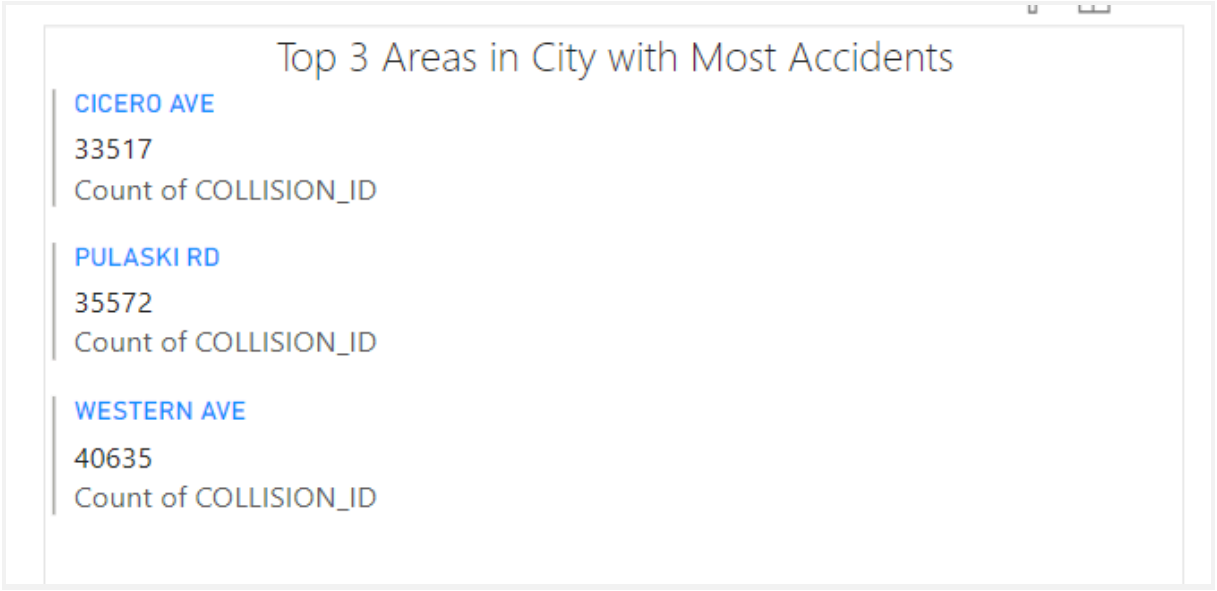
	CITY	AREA	ACCIDENT_COUNT
1	AUSTIN	LAMAR	6286
2	AUSTIN	IH 35	4259
3	AUSTIN	MOPAC	4032
4	CHICAGO	WESTERN AVE	24505
5	CHICAGO	PULASKI RD	21700
6	CHICAGO	CICERO AVE	20190
7	MONTGOMERY	GEORGIA AVE	5600
8	MONTGOMERY	NEW HAMPSHIRE AVE	3542
9	MONTGOMERY	FREDERICK RD	2999
10	NEW YORK CITY	BROADWAY	20561
11	NEW YORK CITY	BELT PARKWAY	18996
12	NEW YORK CITY	ATLANTIC AVENUE	18026

Visualization:

For Austin:

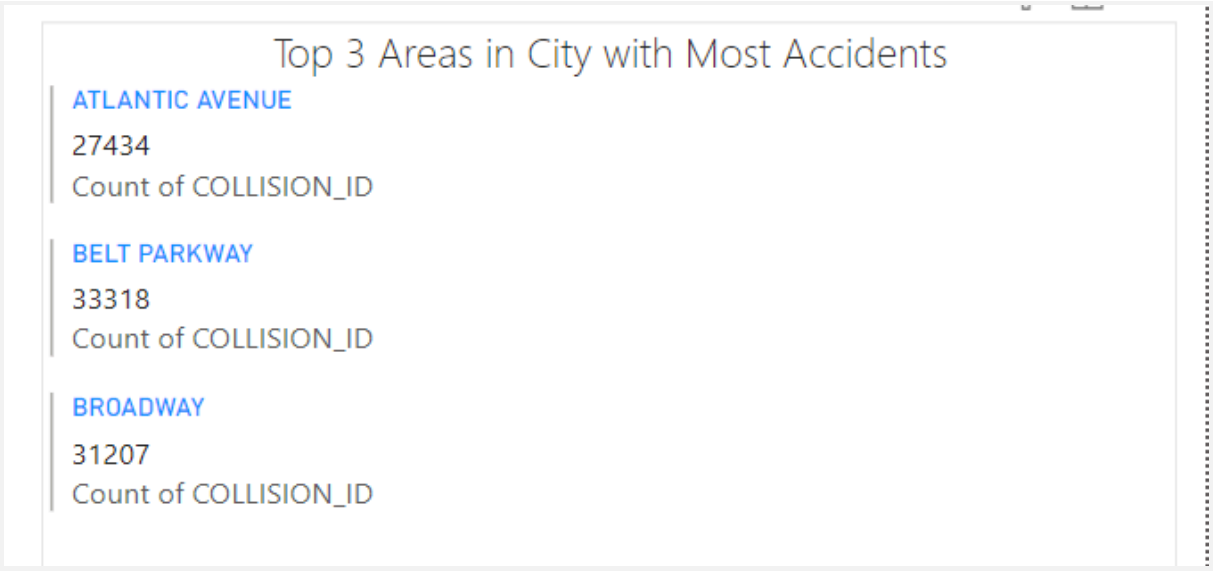


For Chicago:



For Montgomery:

For NYC:



3. How many accidents resulted in just injuries

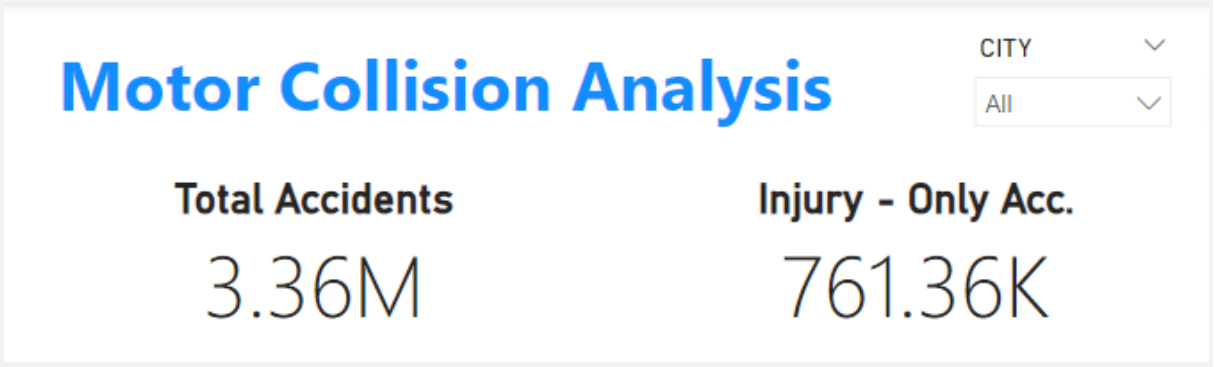
SQL Validation:

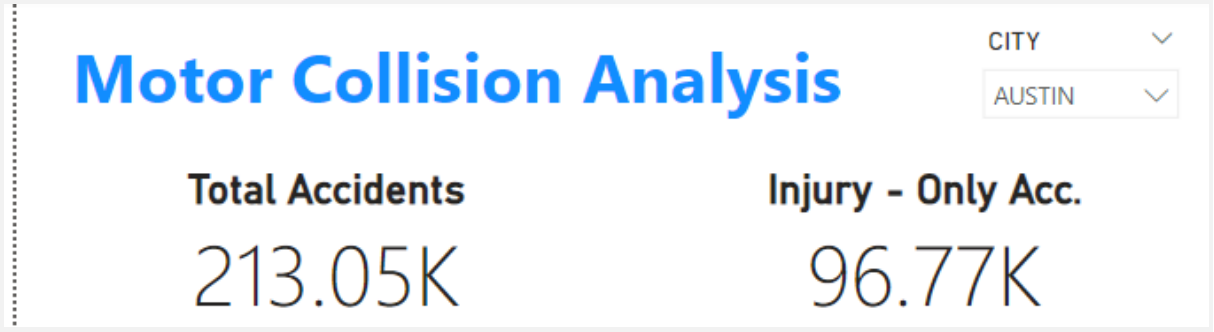
```sql
--- Accidents resulted in just injuries
-- Overall and by city report for accidents resulting in just injuries
WITH ValidAccidents AS (
    SELECT
        da.City,
        COUNT(DISTINCT fc.Collision_ID) AS Injury_Only_Count
    FROM
        FACT_COLLISION fc
    JOIN
        DIM_ADDRESS da ON fc.Address_SK = da.Address_SK
    WHERE
        -- da.City NOT IN ('NA', '') AND da.Street_Name NOT IN ('NA', '')
        fc.Is_Injury_Not_Fatal = 1  -- Only include accidents with injuries but no fatalities
        AND fc.Total_Fatal_Count = 0
    GROUP BY
        da.City
)
-- Combine city-level and overall summary
SELECT
    'Overall' AS City,
    SUM(Injury_Only_Count) AS Injury_Only_Count
```

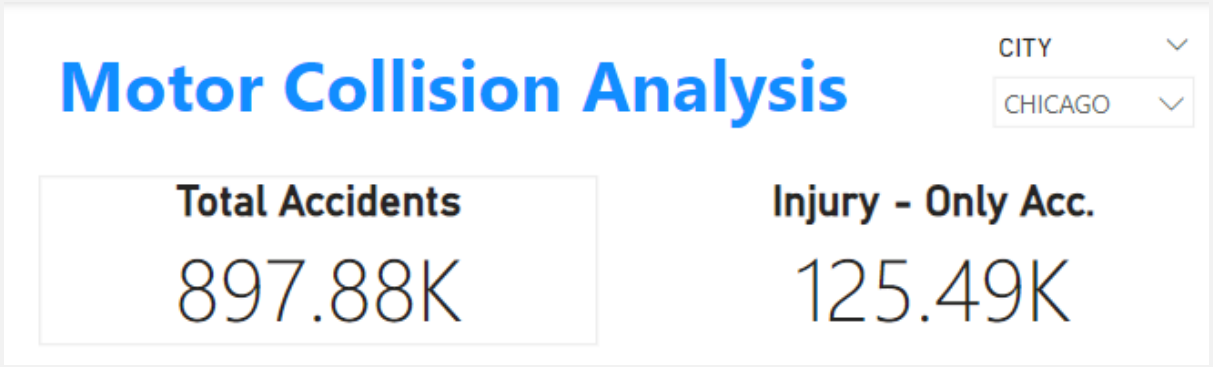
Results | Chart

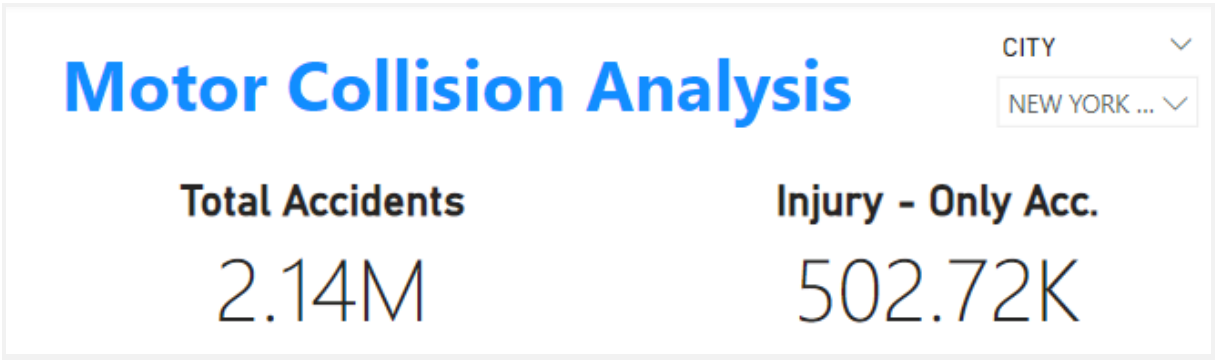
	CITY	INJURY_ONLY_COUNT
1	AUSTIN	96774
2	CHICAGO	125152
3	MONTGOMERY	36371
4	NEW YORK CITY	501911
5	Overall	760208

Visualization:









4. How often are pedestrians involved in accidents (for Chicago)?

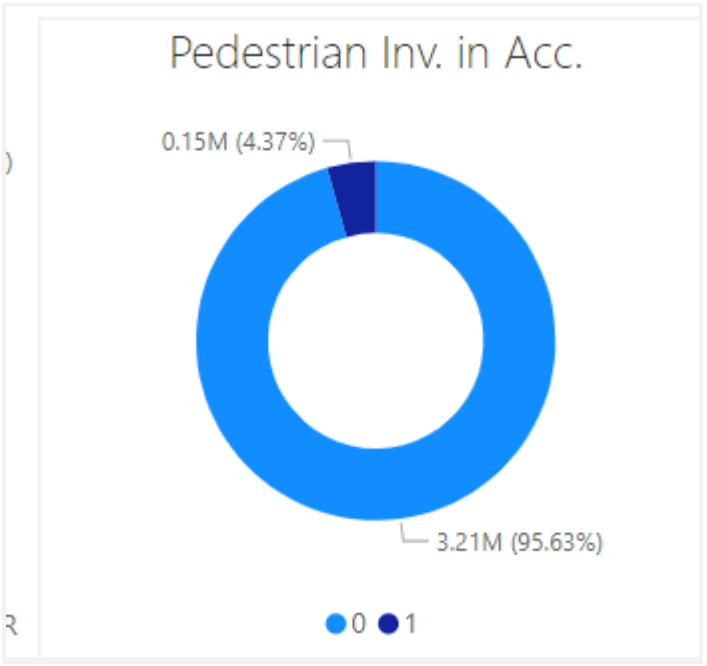
SQL:

```
----How often are pedestrians involved in accidents
SELECT COUNT(DISTINCT(COLLISION_ID)), IS_PEDESTRIAN FROM FACT_COLLISION GROUP BY IS_PEDESTRIAN;

--- Seasonality analysis
```

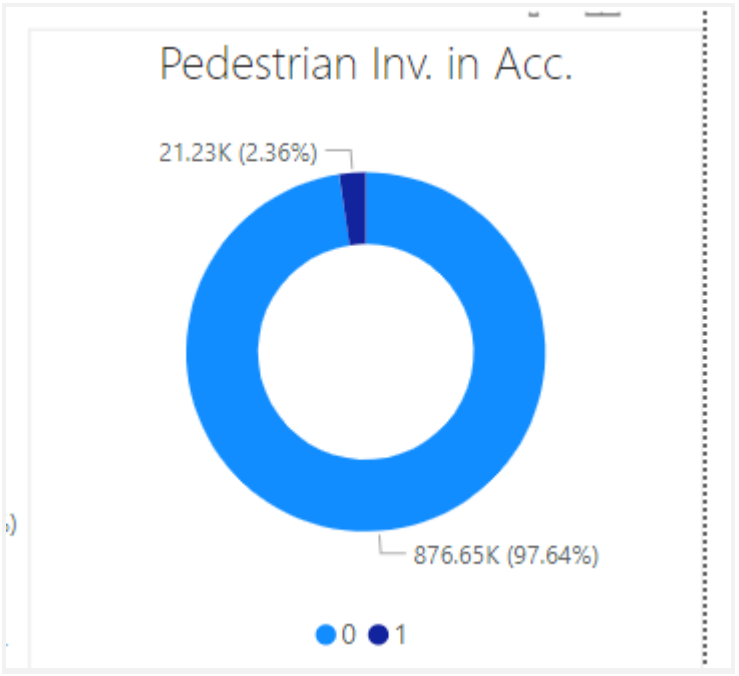
Results | Chart

	COUNT(DISTINCT(COLLISION_ID))	IS_PEDESTRIAN
1	3210251	0
2	146733	1

Overall



For Chicago:




```sql
SELECT COUNT(DISTINCT COLLISION_ID), IS_PEDESTRIAN, a.CITY
FROM FACT_COLLISION f
JOIN DIM_ADDRESS a ON f.ADDRESS_SK = a.ADDRESS_SK
GROUP BY IS_PEDESTRIAN, a.CITY
having city = 'CHICAGO';
```

Results | Chart

	COUNT(DISTINCT COLLISION_ID)	IS_PEDESTRIAN	CITY
1	876646	0	CHICAGO
2	21234	1	CHICAGO

5. When do most accidents happen?

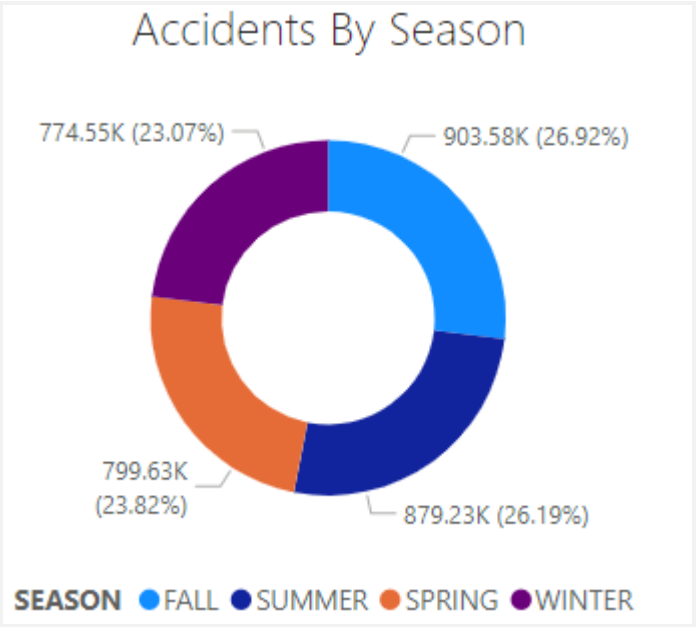
SQL:

```sql
--- Seasonality analysis
SELECT
    d.Season,
    COUNT(DISTINCT f.COLLISION_ID) as total_accidents,
    ROUND((COUNT(DISTINCT f.COLLISION_ID) * 100.0) /
          SUM(COUNT(DISTINCT f.COLLISION_ID)) OVER (), 2) as percentage
FROM FACT_COLLISION f
JOIN DIM_DATE d ON f.Date_SK = d.Date_SK
WHERE f.COLLISION_ID IS NOT NULL
GROUP BY d.Season
ORDER BY total_accidents DESC;
```

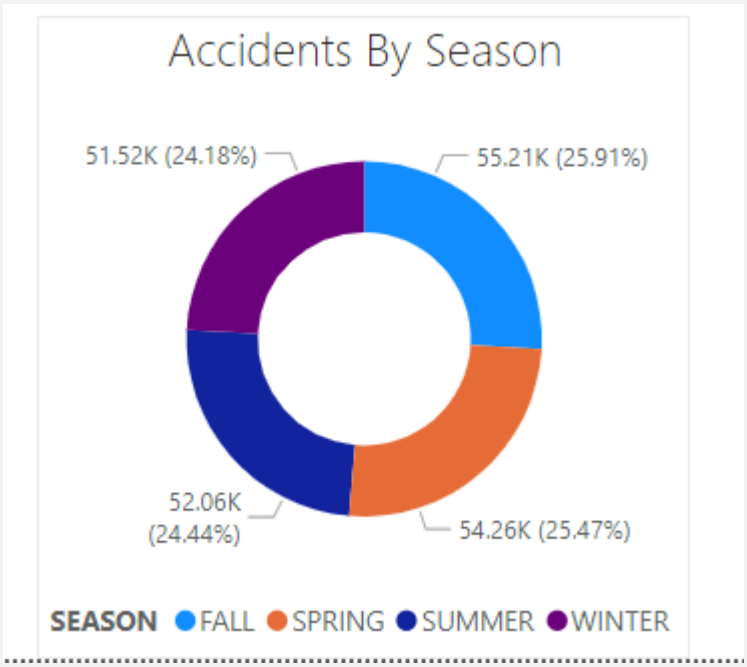
Results | Chart

	SEASON	TOTAL_ACCIDENTS	PERCENTAGE
1	FALL	903578	26.92
2	SUMMER	879227	26.19
3	SPRING	799627	23.82
4	WINTER	774552	23.07

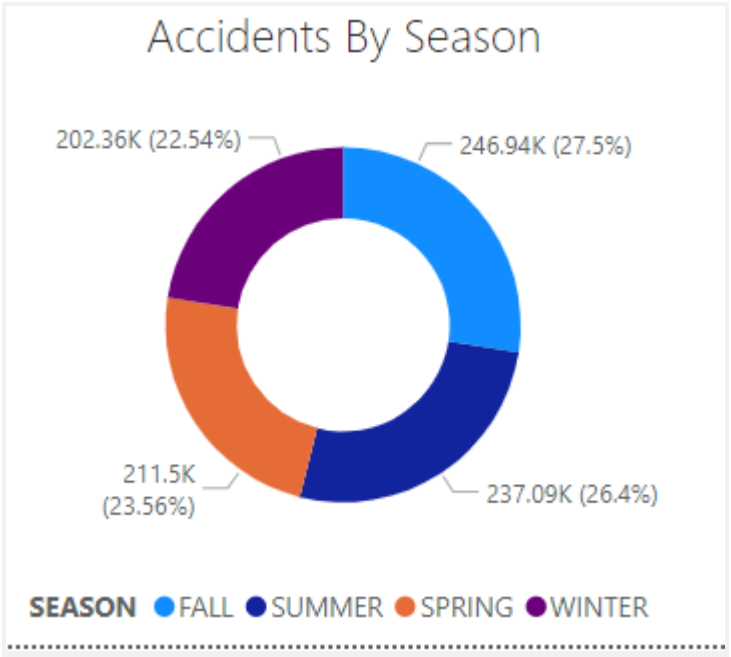
OverallL



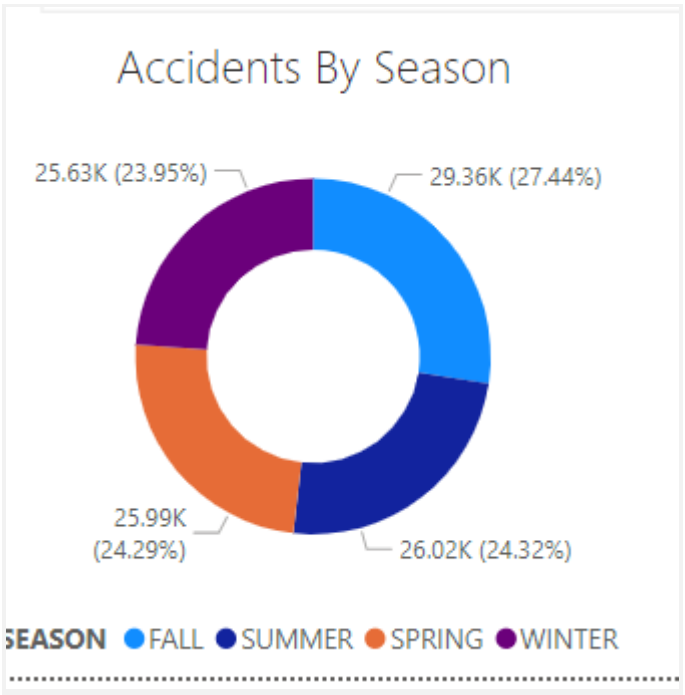
By city: Austin



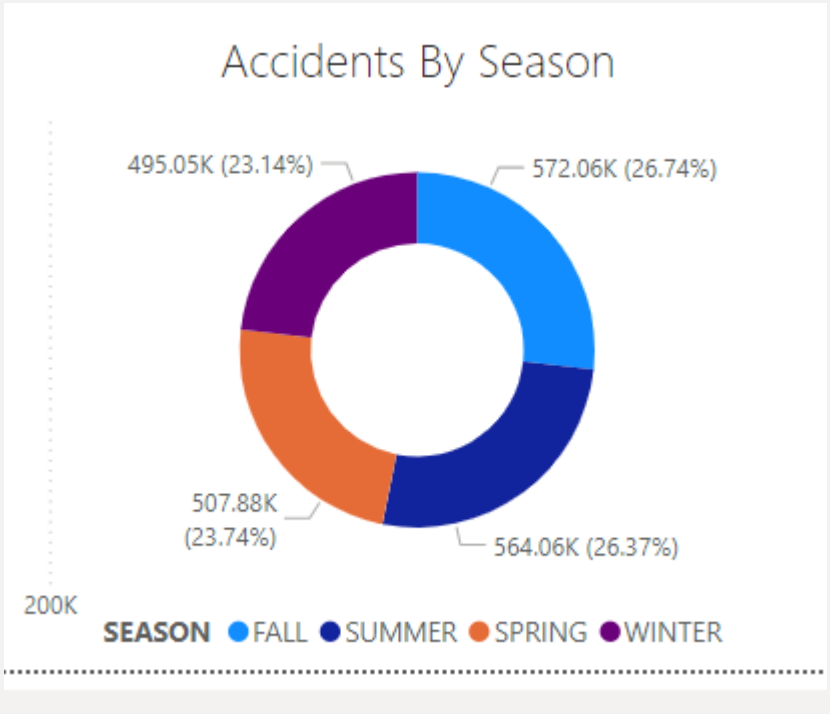
By city: Chicago



Montgomery:



New York:



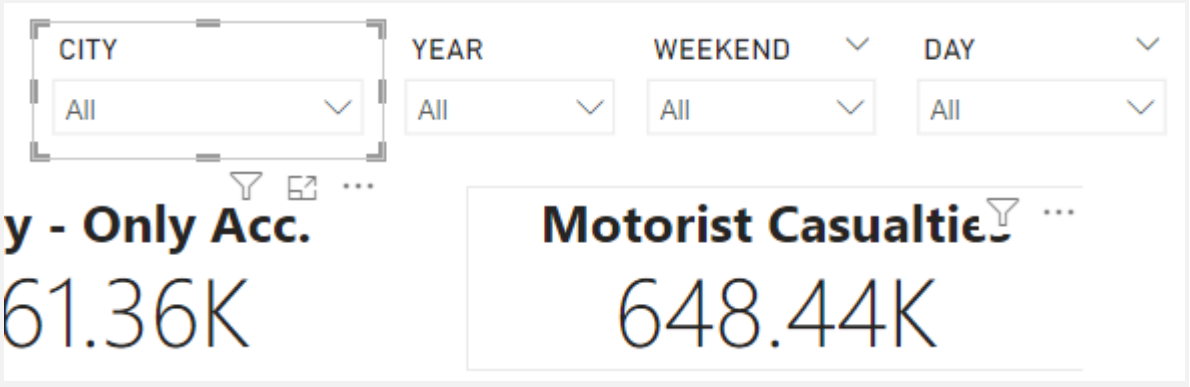
6. How many motorists are injured or killed in accidents (overall and by City)?

```
--Number of motorist casualties overall
SELECT SUM(MOTORIST_DAMAGE)
FROM (SELECT DISTINCT COLLISION_ID, MOTORIST_DAMAGE FROM FACT_COLLISION);
--- Motorist casualties by city
```

Results | Chart

SUM(MOTORIST_DAMAGE)
648441

Overall

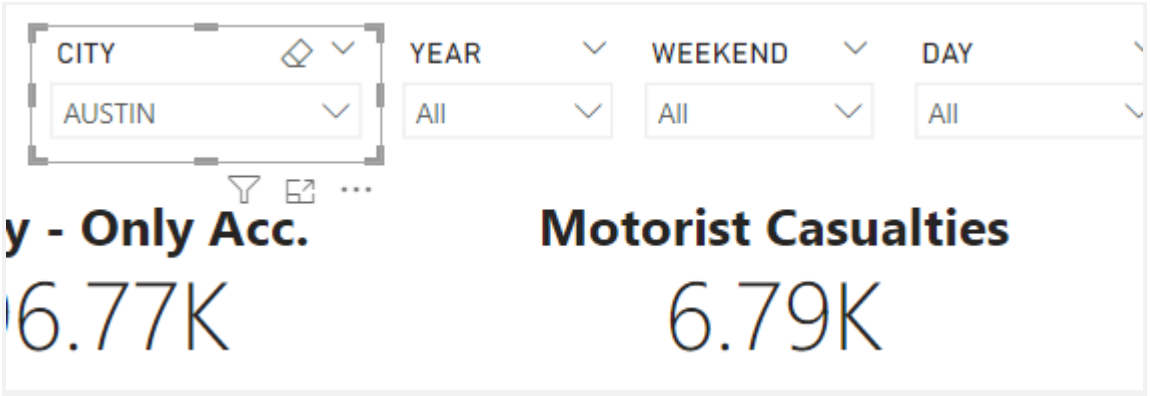


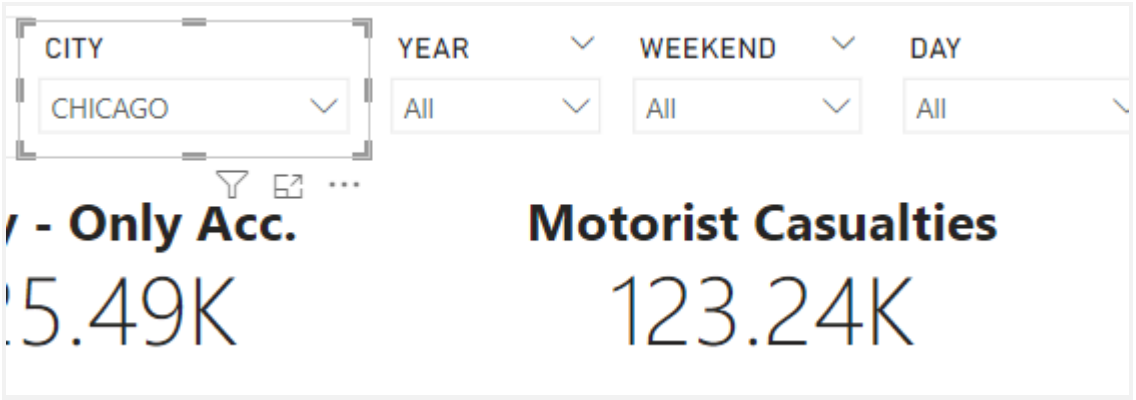
By city:

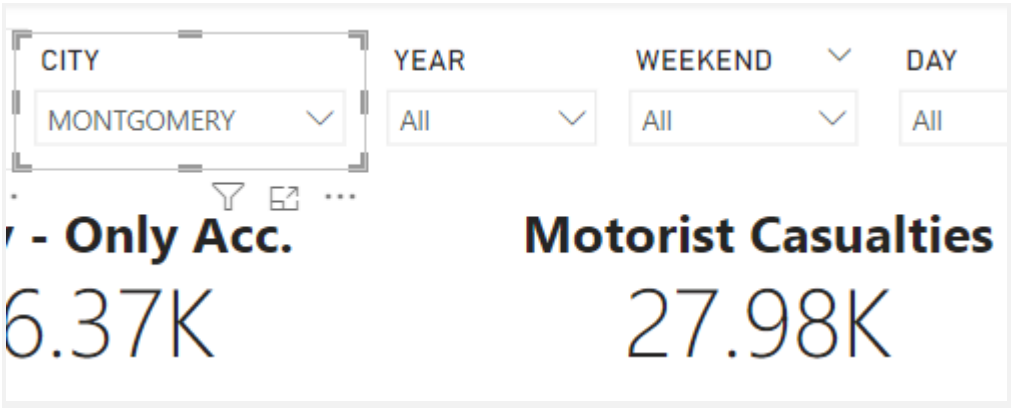
```sql
--- Motorist casualties by city
SELECT SUM(MOTORIST_DAMAGE), address.CITY
FROM (
    SELECT DISTINCT fact.COLLISION_ID, fact.MOTORIST_DAMAGE, fact.ADDRESS_SK
    FROM FACT_COLLISION as fact
) as distinct_facts
JOIN DIM_ADDRESS as address ON distinct_facts.ADDRESS_SK = address.ADDRESS_SK
GROUP BY address.CITY;


--Total fatal count
```

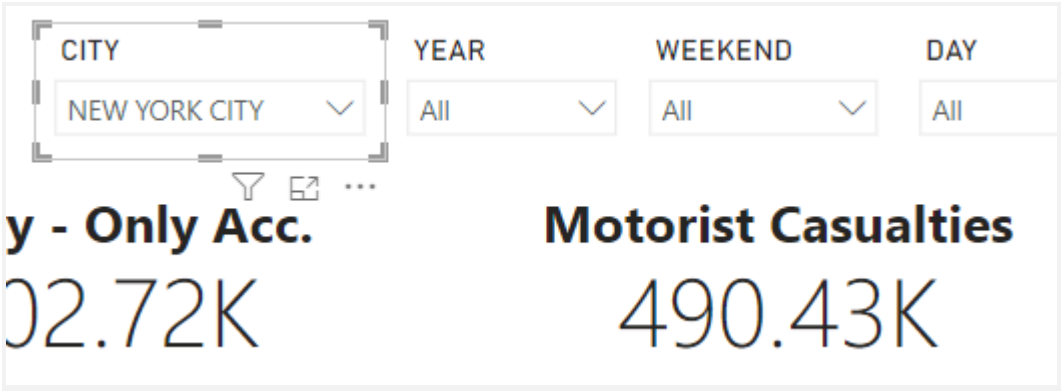
Results | Chart

	SUM(MOTORIST_DAMAGE)	CITY
1	490434	NEW YORK CITY
2	6791	AUSTIN
3	123236	CHICAGO
4	27980	MONTGOMERY









7. Which top 5 areas in the city has the most no.of accidents?

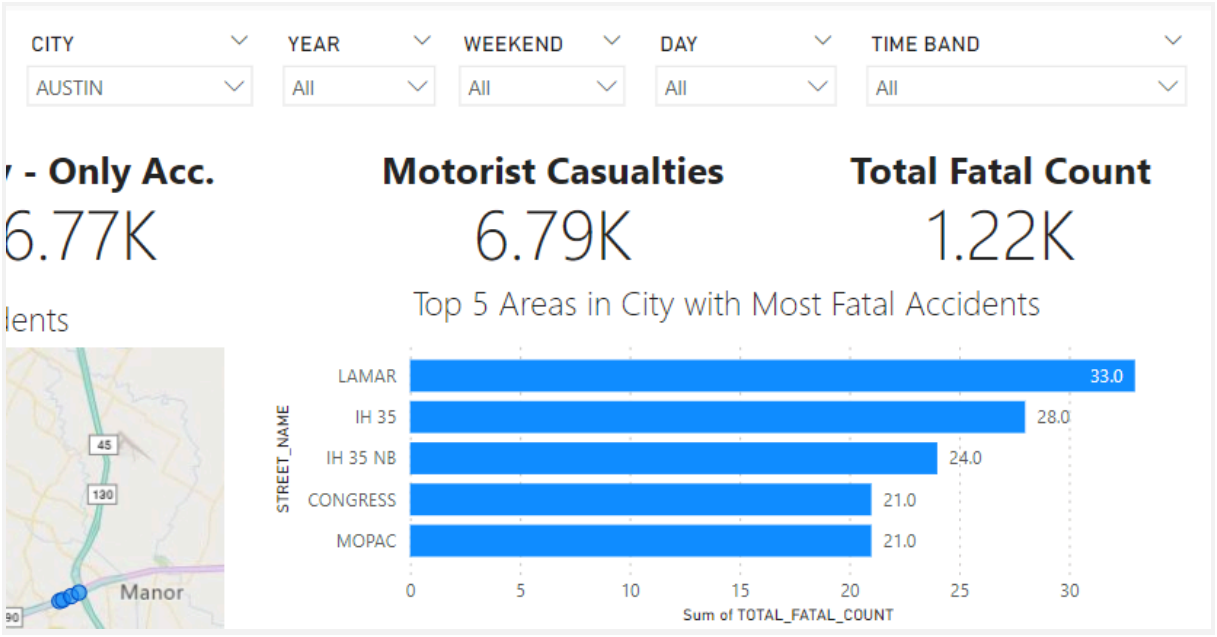
SQL:

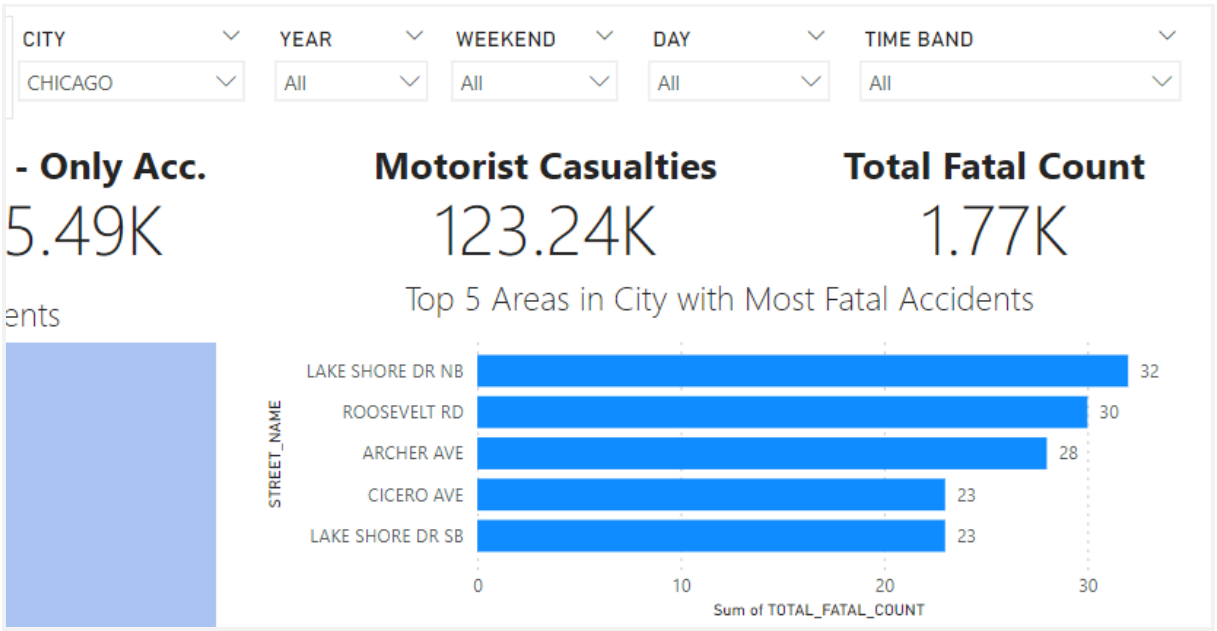
```sql
--Top 5 areas in the city with most fatal no.of accidents
WITH RankedAreas AS (
    SELECT
        a.City AS City,
        a.Street_Name AS Street,
        SUM(f.Total_Fatal_Count) AS Total_Fatal_Accidents,
        RANK() OVER (PARTITION BY a.City ORDER BY SUM(f.Total_Fatal_Count) DESC) AS Rank
    FROM
        FACT_COLLISION f
    JOIN
        DIM_ADDRESS a ON f.Address_SK = a.Address_SK
    WHERE
        a.Street_Name IS NOT NULL AND a.Street_Name != 'NA'
    GROUP BY
```

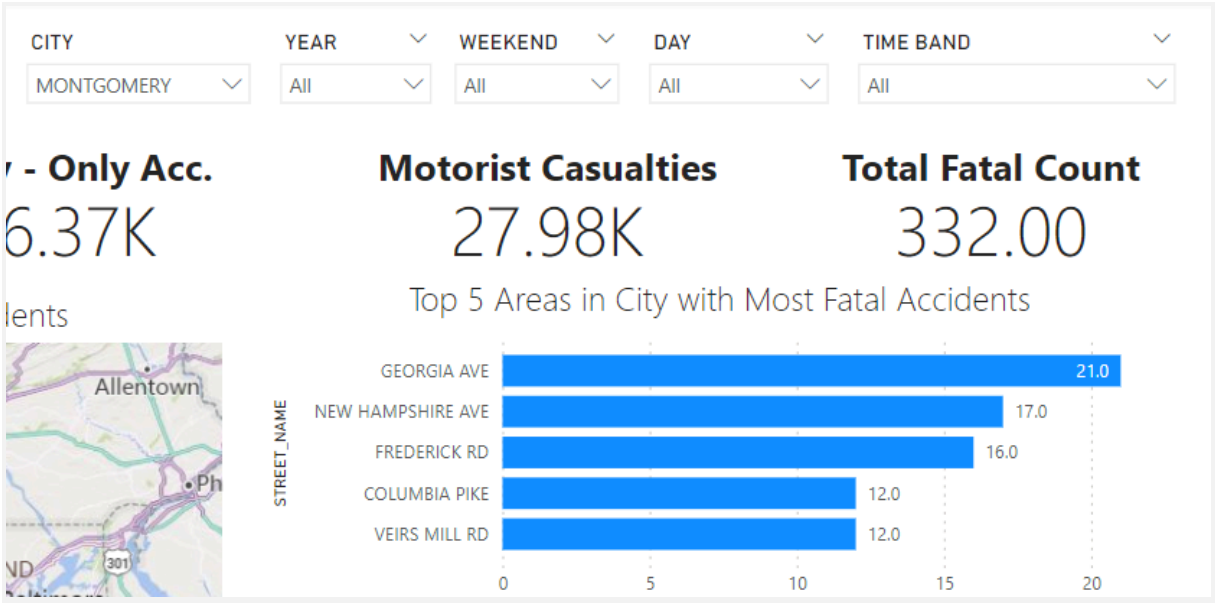
Results | Chart

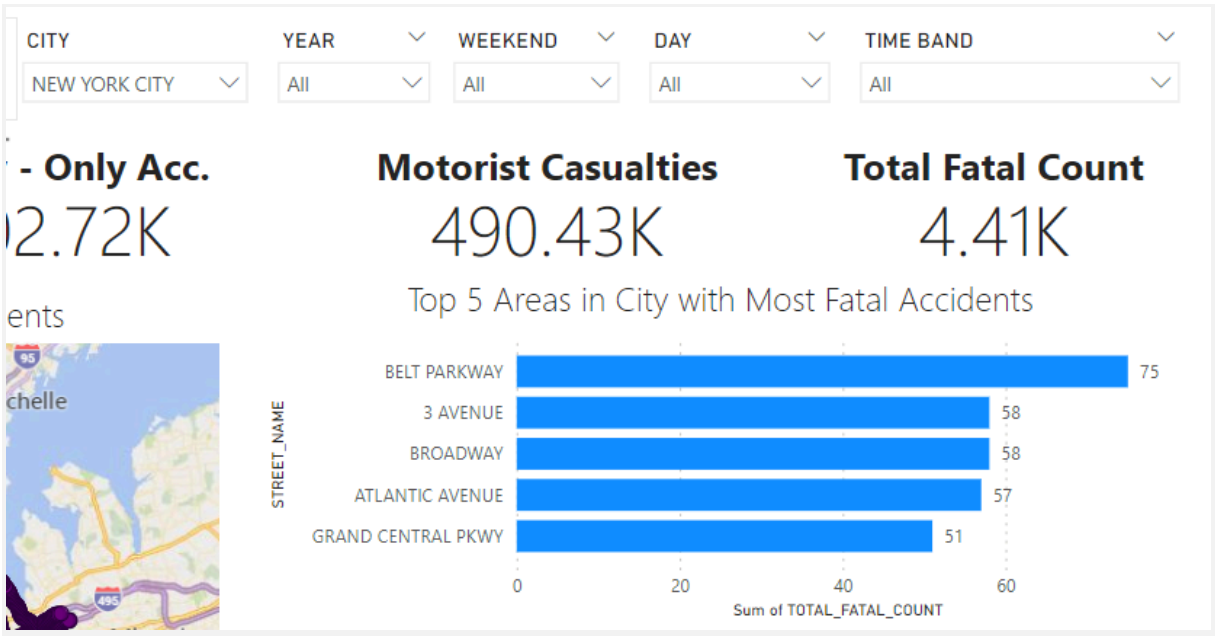
	CITY	STREET	TOTAL_FATAL_ACCIDENTS
1	AUSTIN	LAMAR	33
2	AUSTIN	IH 35	28
3	AUSTIN	IH 35 NB	24
4	AUSTIN	MOPAC	21
5	AUSTIN	CONGRESS	21
6	CHICAGO	LAKE SHORE DR NB	32
7	CHICAGO	ROOSEVELT RD	30
8	CHICAGO	ARCHER AVE	28
9	CHICAGO	CICERO AVE	23
10	CHICAGO	LAKE SHORE DR SB	23
11	MONTGOMERY	GEORGIA AVE	21
12	MONTGOMERY	NEW HAMPSHIRE AVE	17
13	MONTGOMERY	FREDERICK RD	16
16	NEW YORK CITY	BELT PARKWAY	75
17	NEW YORK CITY	BROADWAY	58
18	NEW YORK CITY	3 AVENUE	58
19	NEW YORK CITY	ATLANTIC AVENUE	57
20	NEW YORK CITY	GRAND CENTRAL PKWY	51

Visualization:


CITY
AUSTIN
YEAR
All
WEEKEND
All
DAY
All
TIME BAND
All
- Only Acc.
6.77K
lents
Motorist Casualties
6.79K
Total Fatal Count
1.22K
Top 5 Areas in City with Most Fatal Accidents
45
130
Manor
STREET_NAME
LAMAR
IH 35
IH 35 NB
CONGRESS
MOPAC
33.0
28.0
24.0
21.0
21.0
0
5
10
15
20
25
30
Sum of TOTAL_FATAL_COUNT


CITY
CHICAGO
YEAR
All
WEEKEND
All
DAY
All
TIME BAND
All
- Only Acc.
5.49K
ents
Motorist Casualties
123.24K
Total Fatal Count
1.77K
Top 5 Areas in City with Most Fatal Accidents
STREET_NAME
LAKE SHORE DR NB
ROOSEVELT RD
ARCHER AVE
CICERO AVE
LAKE SHORE DR SB
32
30
28
23
23
0
10
20
30
Sum of TOTAL_FATAL_COUNT





8. Time based analysis of the accident:
   Time of the day, Day of the week, Weekend or Weekdays

SQL:

```sql
---- Based on time of a day
SELECT
    t.Time_Band_Label AS time_of_day,
    COUNT(DISTINCT f.COLLISION_ID) AS total_accidents,
    ROUND((COUNT(DISTINCT f.COLLISION_ID) * 100.0) /
          SUM(COUNT(DISTINCT f.COLLISION_ID)) OVER (), 2) as percentage
FROM FACT_COLLISION f
JOIN DIM_TIME t ON f.Time_SK = t.Time_SK
GROUP BY t.Time_Band_Label
ORDER BY total_accidents DESC;
```

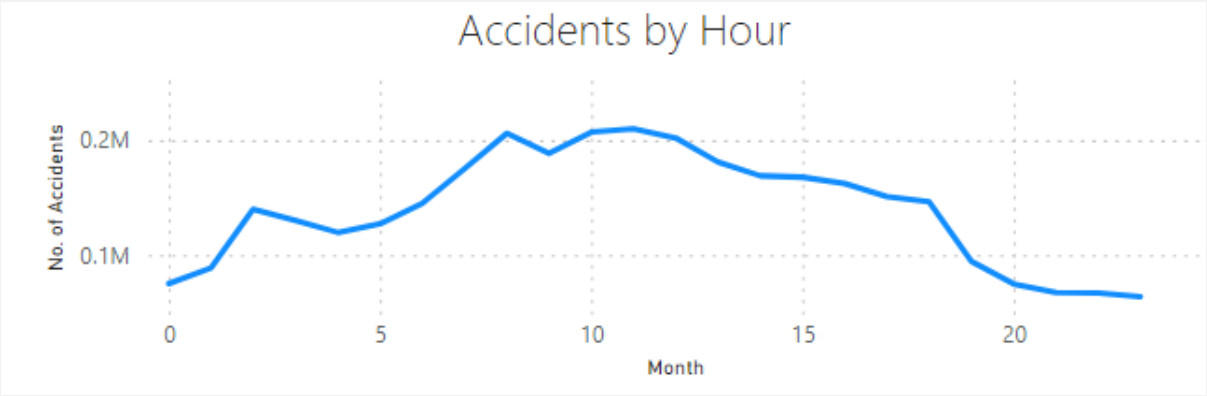
Results | Chart

	TIME_OF_DAY	TOTAL_ACCIDENTS	PERCENTAGE
1	MIDDAY ACTIVITY	969022	28.87
2	EVENING COMMUTE	721374	21.49
3	MORNING COMMUTE	714626	21.29
4	OVERNIGHT	679953	20.25
5	NIGHTTIME	272009	8.10

```sql
SELECT
    t.HOUR AS time_of_day,
    COUNT(DISTINCT f.COLLISION_ID) AS total_accidents,
    ROUND((COUNT(DISTINCT f.COLLISION_ID) * 100.0) /
          SUM(COUNT(DISTINCT f.COLLISION_ID)) OVER (), 2) as percentage
FROM FACT_COLLISION f
JOIN DIM_TIME t ON f.Time_SK = t.Time_SK
GROUP BY t.HOUR
ORDER BY total_accidents DESC;
```

Results | Chart

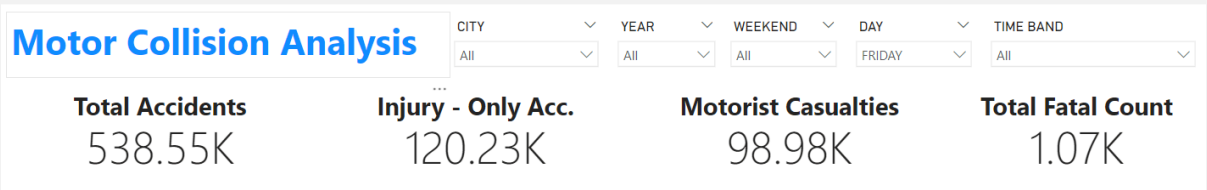
	TIME_OF_DAY	TOTAL_ACCIDENTS	PERCENTAGE
1	11	210084	6.26
2	10	207055	6.17
3	8	205969	6.14
4	12	202004	6.02
5	9	188614	5.62
6	13	180954	5.39
7	7	175064	5.21
8	14	168925	5.03
9	15	167716	5.00
10	16	162217	4.83
11	17	150876	4.49
12	18	146182	4.35
13	6	144979	4.32

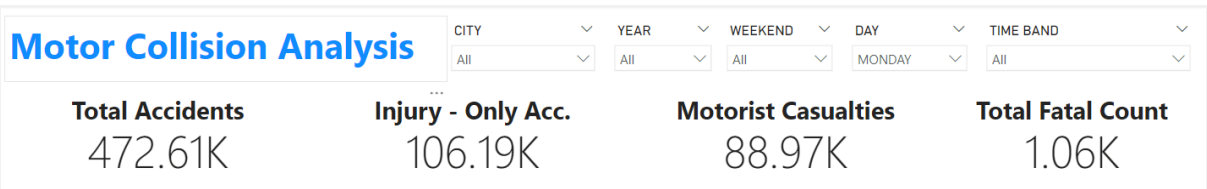



```sql
SELECT
    d.Day AS day_of_week,
    COUNT(DISTINCT f.COLLISION_ID) AS total_accidents,
    ROUND((COUNT(DISTINCT f.COLLISION_ID) * 100.0) /
          SUM(COUNT(DISTINCT f.COLLISION_ID)) OVER (), 2) as percentage
FROM FACT_COLLISION f
JOIN DIM_DATE d ON f.Date_SK = d.Date_SK
GROUP BY d.Day
ORDER BY total_accidents DESC;

---Weekdays or weekend Analysis
```

Results | Chart

	DAY_OF_WEEK	TOTAL_ACCIDENTS	PERCENTAGE
1	FRIDAY	538545	16.04
2	THURSDAY	496641	14.79
3	TUESDAY	489444	14.58
4	WEDNESDAY	486407	14.49
5	MONDAY	472610	14.08
6	SATURDAY	465736	13.87
7	SUNDAY	407601	12.14

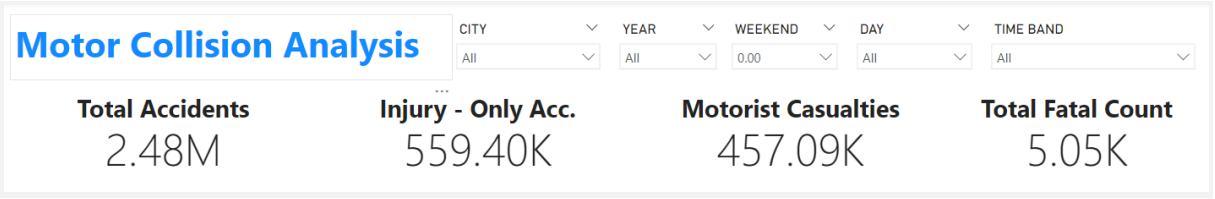




```sql
---Weekdays or weekend Analysis
SELECT
    CASE WHEN d.Is_Weekend = 1 THEN 'Weekend' ELSE 'Weekday' END AS day_type,
    COUNT(DISTINCT f.COLLISION_ID) AS total_accidents,
    ROUND((COUNT(DISTINCT f.COLLISION_ID) * 100.0) /
          SUM(COUNT(DISTINCT f.COLLISION_ID)) OVER (), 2) as percentage
FROM FACT_COLLISION f
JOIN DIM_DATE d ON f.Date_SK = d.Date_SK
GROUP BY day_type
```

Results | Chart

	DAY_TYPE	TOTAL_ACCIDENTS	PERCENTAGE
1	Weekday	2483647	73.98
2	Weekend	873337	26.02



9. Fatality Analysis:

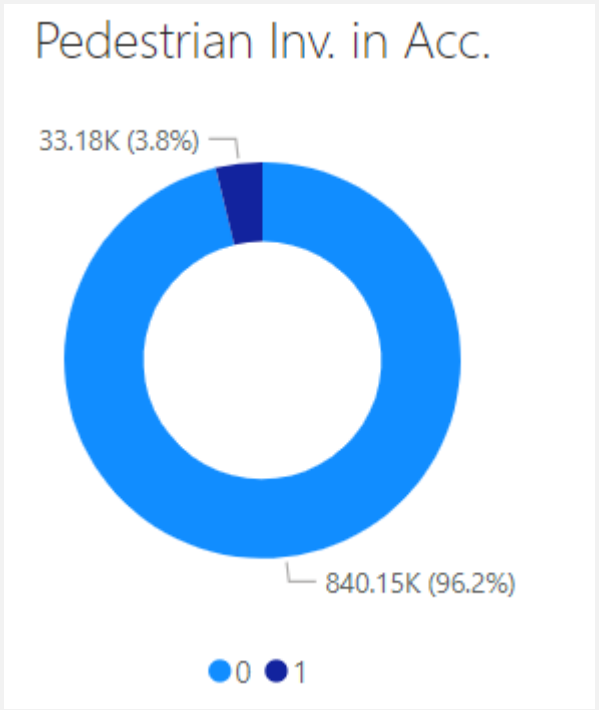
Ans:

SQL:

```sql
SELECT COUNT(DISTINCT(COLLISION_ID)), IS_PEDESTRIAN, from fact_collision
GROUP BY IS_PEDESTRIAN;
```

Results | Chart

	COUNT(DISTINCT(COLLISION_ID))	IS_PEDESTRIAN
1	3210251	0
2	146733	1



10. What are the most common factors involved in accidents?

Ans:

SQL:

```sql
-- Most common factors or contributions
SELECT
    c.DESCRIPTION AS contribution_description,
    c.NORMALIZED_DESCRIPTION AS normalized_factor,
    COUNT(DISTINCT f.COLLISION_ID) as accident_count,
    ROUND((COUNT(DISTINCT f.COLLISION_ID) * 100.0) /
          SUM(COUNT(DISTINCT f.COLLISION_ID)) OVER (), 2) as percentage
FROM FACT_COLLISION f
JOIN DIM_CONTRIBUTION c ON f.Contribution_SK = c.Contribution_SK
WHERE c.DESCRIPTION IS NOT NULL
    AND c.DESCRIPTION != 'NA'
```

Results | Chart

	CONTRIBUTION_DESCRIPTION	NORMALIZED_FACTOR	ACCIDENT_COUNT	PERCENTAGE
1	Unable to Determine	UNABLE TO DETERMINE	467802	20.91
2	other	NOT APPLICABLE	373277	16.69
3	Followed Too Closely	FOLLOWING TOO CLOSELY	225583	10.08
4	Failed to Drive in Single Lane	FAILURE TO YIELD RIGHT-OF-WAY	138158	6.18
5	FAILING TO YIELD RIGHT-OF-WAY	FAILING TO YIELD RIGHT-OF-WAY	114790	5.13
6	Backed without Safety	BACKING UNSAFELY	84489	3.78
7	Unsafe Speed	FAILING TO REDUCE SPEED TO AVOID CRASH	68099	3.04
8	Passed in No Passing Lane	PASSING OR LANE USAGE IMPROPER	67808	3.03
9	Passed on Shoulder	PASSING TOO CLOSELY	59408	2.66
10	Turned Improperly (Cut Corner on Left)	TURNING IMPROPERLY	58540	2.62
11	DRIVING SKILLS/KNOWLEDGE/EXPERIENCE	DRIVING SKILLS/KNOWLEDGE/EXPERIENCE	54348	2.43
12	Overtake and Pass Insufficient Clearance	IMPROPER OVERTAKING/PASSING	52291	2.34
13	Turned Improperly (Wrong Lane)	UNSAFE LANE CHANGING	46468	2.08
14	Failed to Drive in Single Lane	IMPROPER LANE USAGE	40969	1.83