UFO Sightings - SQL/Tableau

As my first data analysis project, I wanted to explore a light-hearted topic and found a UFO sightings dataset on Kaggle that instantly grabbed my attention.

The Data

The dataset has 80332 rows with columns for the date/time, city, state/province, country, UFO shape, length of the encounter (seconds), described length of encounter, description of the encounter, date documented, and latitude/longitude.

From the information shown on the table I decided to ask the question of when and where are you most likely to see a UFO?

Cleaning

First thing that I noticed for cleaning was the date/time column of the data. I knew I wanted to be able to reference date and time separately in my query, so I used the CONVERT statement to convert the datetime data type into just date and the ALTER TABLE and UPDATE statements to create a new column for the converted data and propagate the new data into that column. I repeated for ‘time’ as well.

```
SELECT Date_time, CONVERT(Date, Date_time)
FROM UFO_sightings.dbo.ufo_sighting_data;

ALTER TABLE ufo_sighting_data
ADD Sighting_Date Date;

UPDATE ufo_sighting_data
SET  Sighting_Date = CONVERT(Date, Date_time);

ALTER TABLE ufo_sighting_data
ADD Sighting_Time Time;

UPDATE ufo_sighting_data
SET  Sighting_Time = CONVERT(Time, Date_time);
```

I used the DROP COLUMN statement to drop two unnecessary columns - the original “Date_time” column, and the “described_duration_of_encounter” column.

I then had a look at the blank cells and noticed a lot of cells in the “country” column were blank, but their corresponding “state/province” cells had data. In an attempt to conserve as much data as possible, I created a list of abbreviations for Canadian provinces and US states and used the UPDATE, SET, and WHERE IN clauses to fill the missing country data in those cells.

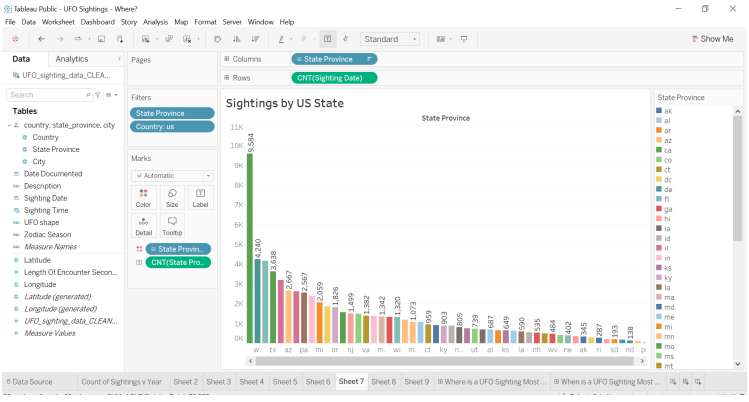
```
UPDATE ufo_sighting_data
SET country = 'ca'
WHERE "state_province" IN ('NL', 'PE', 'NS', 'NB', 'QC', 'ON', 'MB', 'SK', 'AB', 'BC', 'YT', 'NT', 'NU');

UPDATE ufo_sighting_data
SET country = 'us'
WHERE "state_province" IN ('AL', 'AK', 'AZ', 'AR', 'AS', 'CA', 'CO', 'CT', 'DE', 'DC', 'FL', 'GA', 'GU', 'HI', 'ID', 'IL

UPDATE ufo_sighting_data
SET country = 'us'
WHERE "state_province" IN ('mn', 'ms', 'mo', 'mt', 'ne', 'nv', 'nh', 'nj', 'nm', 'ny', 'nc', 'nd');
```

I thought it would be interesting to see in which zodiac seasons UFO sightings were more common, so I added a column to classify each sighting into a zodiac season with the ALTER TABLE statement and then filled that column with complex WHERE statements to determine the zodiac sign of each sighting date.

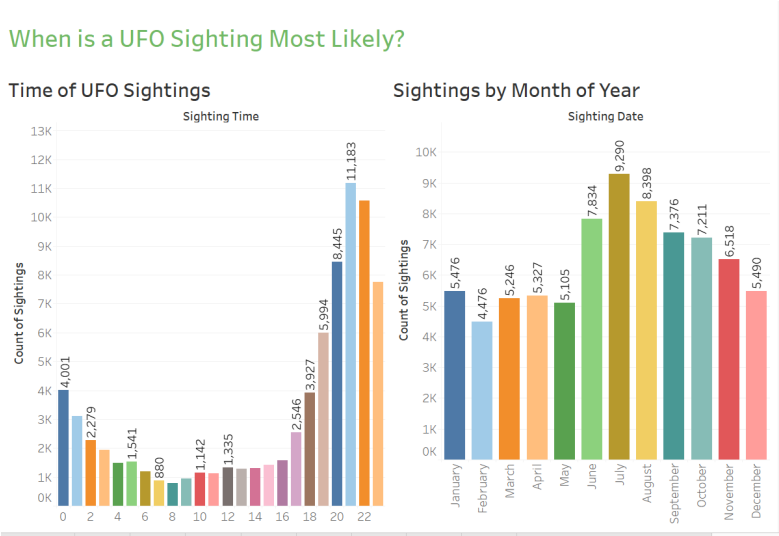
```
--ADD Column 'Zodiac_Season' to classify the season in which the sighting occured

ALTER TABLE UFO_sightings.dbo.ufo_sighting_data
ADD Zodiac_Season varchar(20);

--Use month and day values from datetime Sighting_Date column to classify which zodiac season each sighting occured in

UPDATE UFO_sightings.dbo.ufo_sighting_data
SET Zodiac_Season = 'Aries'
WHERE MONTH(Sighting_Date) = '03' AND DAY(Sighting_Date) BETWEEN '21' AND '31';

UPDATE UFO_sightings.dbo.ufo_sighting_data
SET Zodiac_Season = 'Aries'
WHERE MONTH(Sighting_Date) = '04' AND DAY(Sighting_Date) BETWEEN '1' AND '19';
```

Visualisations

I created visualisations from this data in Tableau - backend view as shown below.



I created a Tableau dashboard showing that UFO sightings are most common between 20:00 and 23:00 in the months of June, July, and August.



And a Tableau dashboard showing that UFO sightings are most common in the US state of California.