PREDICTING CALIFORNIA TRAVEL TRENDS

DSC 630 | Amy Nestingen | Corinne Medeiros

Photo by Bruno Wolff on UnSplash

OVERVIEW

PROBLEM

Many small businesses across California depend heavily on tourist sales. When people travel, they support hotels, restaurants, transportation, and shops.

In 2020, the COVID-19 pandemic has caused a sharp decrease in travel. People are eager to predict when travel will return to normal and tourist shops will regain their sales.

OBJECTIVE

This project aims to predict future travel trends, focusing on the next 6 months for short term planning and the next 2 years for longer term planning.

This information could help small businesses hire back the people they had to lay off and stock up on goods for the return customers.

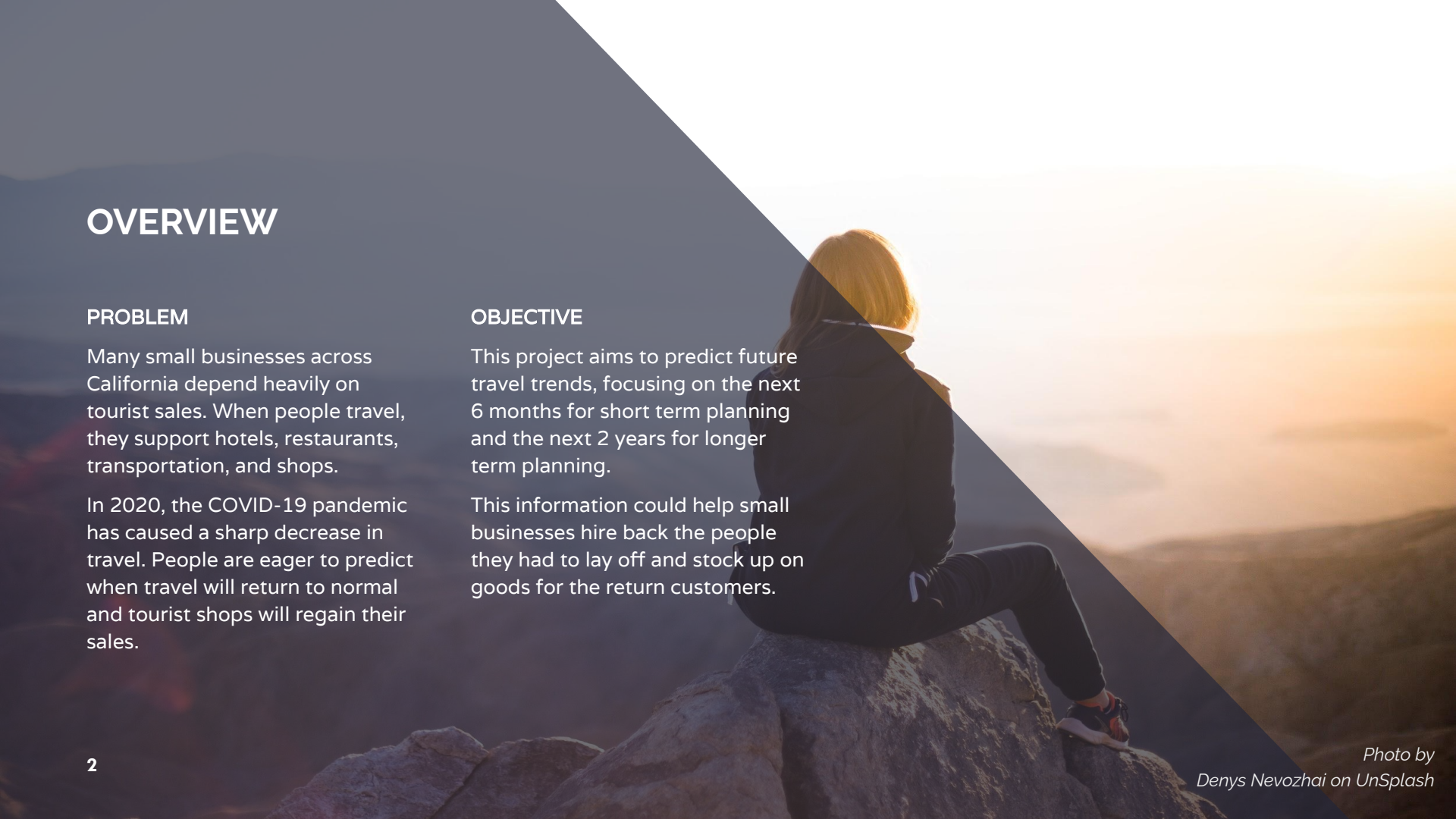


Photo by Denys Nevozhai on UnSplash

1.
DATA EXPLORATION

Data Cleanup & Preliminary Analysis

Photo by Matthew Hamilton on UnSplash

U.S. Department of Transportation
Bureau of Transportation Statistics

OVERVIEW OF TRIP DATA

SOURCE

The Bureau of Transportation Statistics (BTS) provides records from an "anonymized national panel of mobile device data" (BTS.gov., 2020).

Mobility statistics are estimated by the Maryland Transportation Institute and Center for Advanced Transportation Technology Laboratory at the University of Maryland.

ATTRIBUTES

Trips are measured by distance (miles) and are defined as "movements that include a stay of longer than 10 minutes at an anonymized location away from home" (BTS.gov., 2020).

Trips involve all varieties of transportation including cars, trains, public transportation, and flights.

Data source: https://catalog.data.gov/dataset/trips-by-distance

Photo by NASA on UnSplash

DATA CLEANUP

- Removed rows with missing data
- Removed outliers - negative values within population not staying at home and the number of trips taken
- Cleaned up spelling discrepancies in the names of counties
- Limited data to California counties only

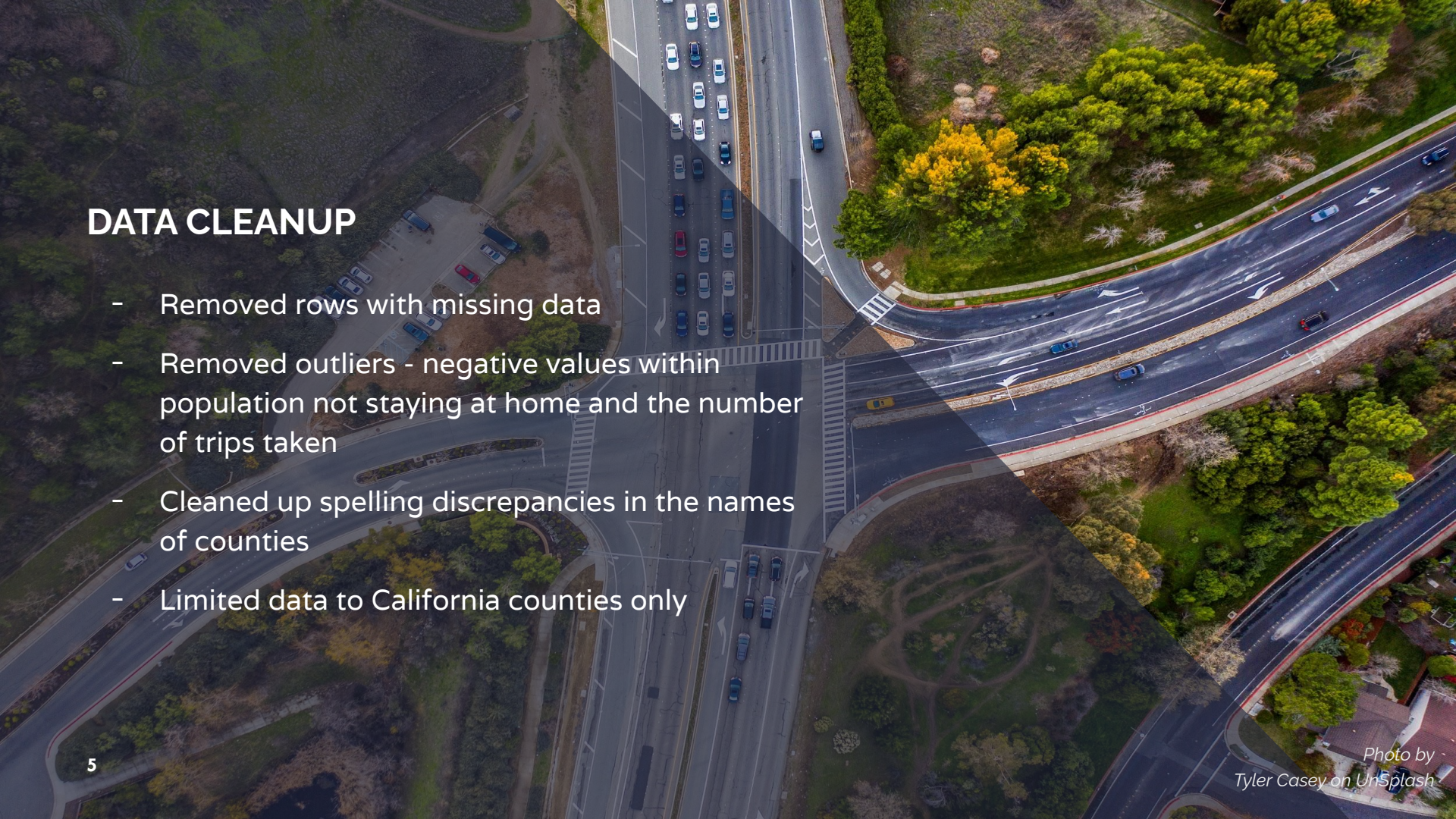
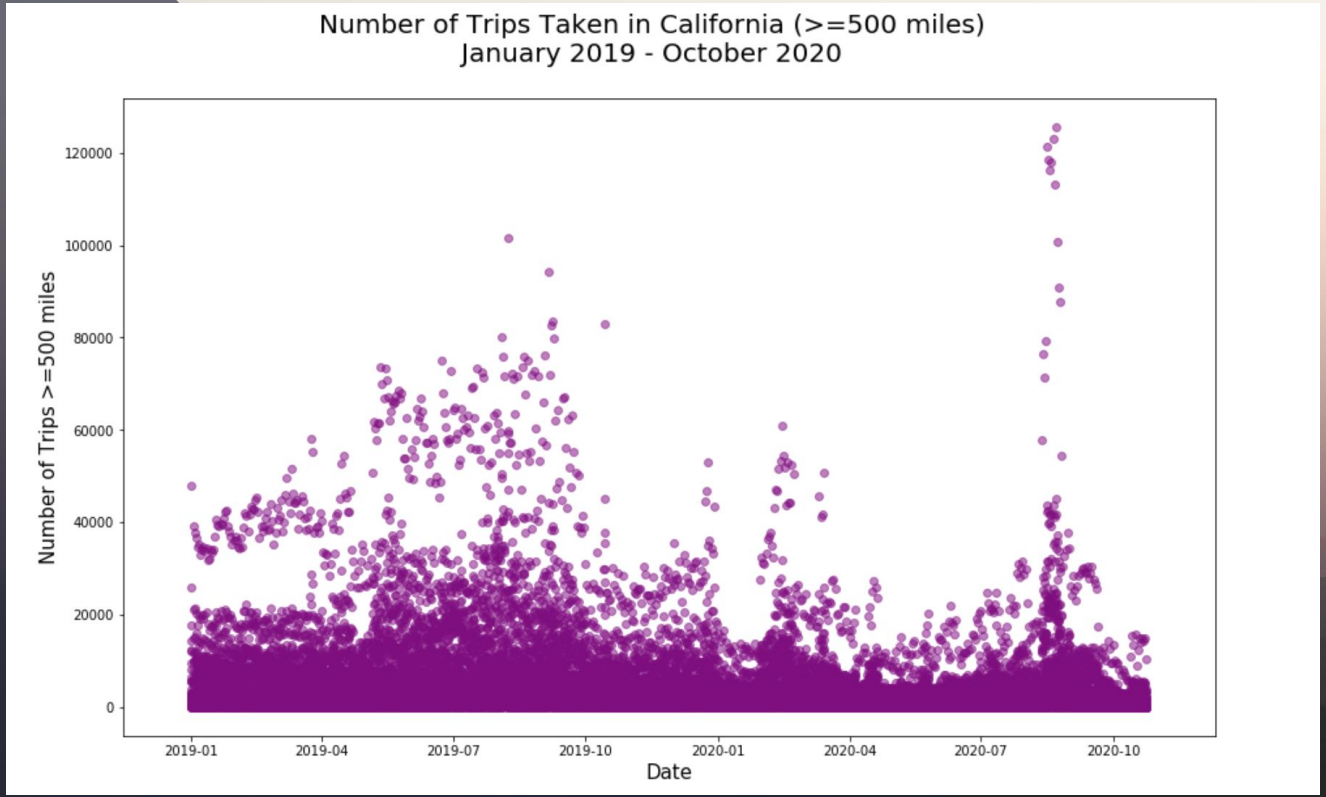


Photo by Tyler Casey on UnSplash



Our data encompasses January 2019 through October 2020.

In the past, we've seen travel decrease during a recession, and now we are seeing that same trend during the COVID-19 pandemic.

Photo by Amanda Frank on UnSplash

OVERVIEW OF COVID-19 DATA

SOURCE

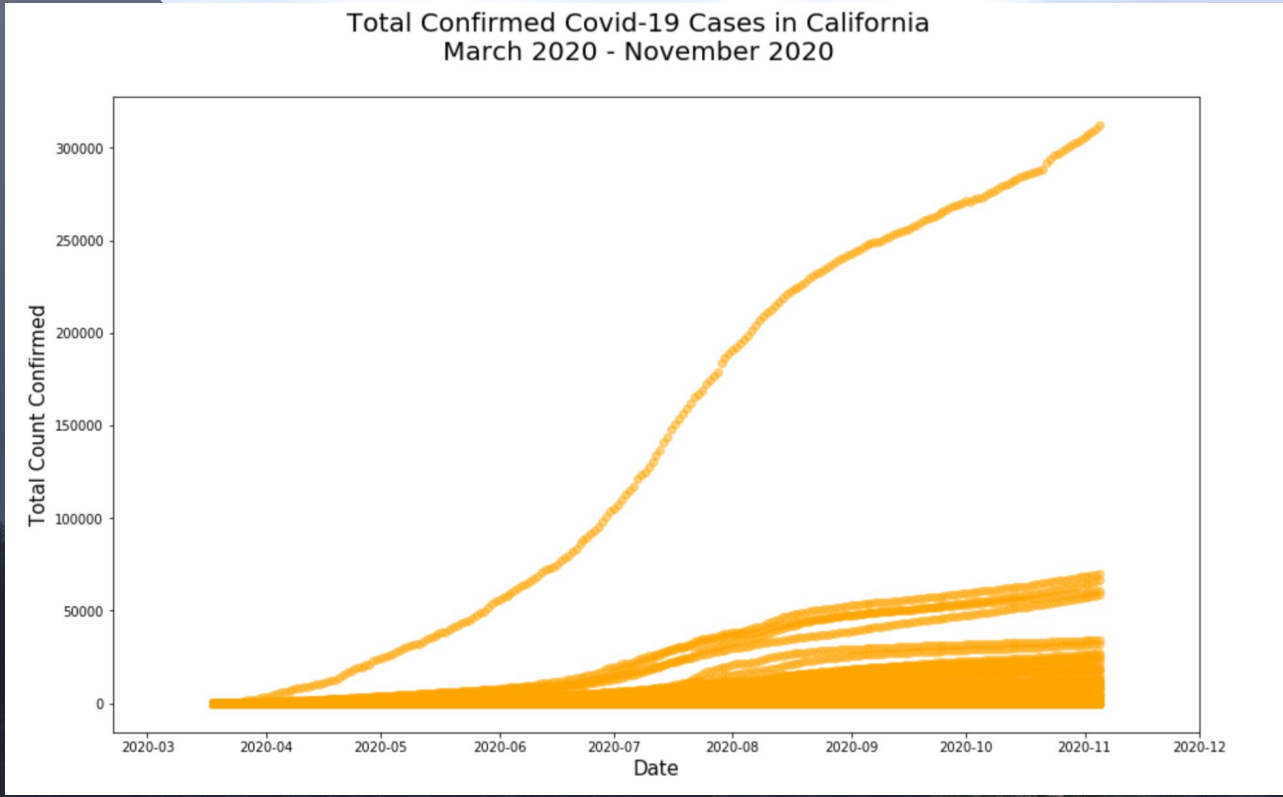
The California Department of Public Health (CDPH) Coronavirus Science Branch, through California Open Data, provides confirmed positive cases and deaths reported by local health departments in California and is updated daily.

ATTRIBUTES

Observations include a cumulative number of laboratory-confirmed positive cases, as well as Covid-related deaths, per county, starting from March 19, 2020.

Data source: https://data.ca.gov/dataset/covid-19-cases/resource/926fd08f-cc91-4828-af38-bd45de97f8c3

Photo by NASA on UnSplash



To compare trends, we supplemented our analysis with additional data reporting positive COVID-19 cases in California.

Photo by Luis Soto on UnSplash

2.
METHODS & RESULTS

Predictive Modeling

Photo by
Pablo Garcia Saldana on UnSplash

Models Tested

- The main criteria that we based the models on are the mean absolute percentage error (MAPE)
- The lower the MAPE, the better the model
- The Models we performed include:
  - Naive
  - Simple Exponential Smoothing
  - ARIMA
  - Holt's Trend
  - TBATS

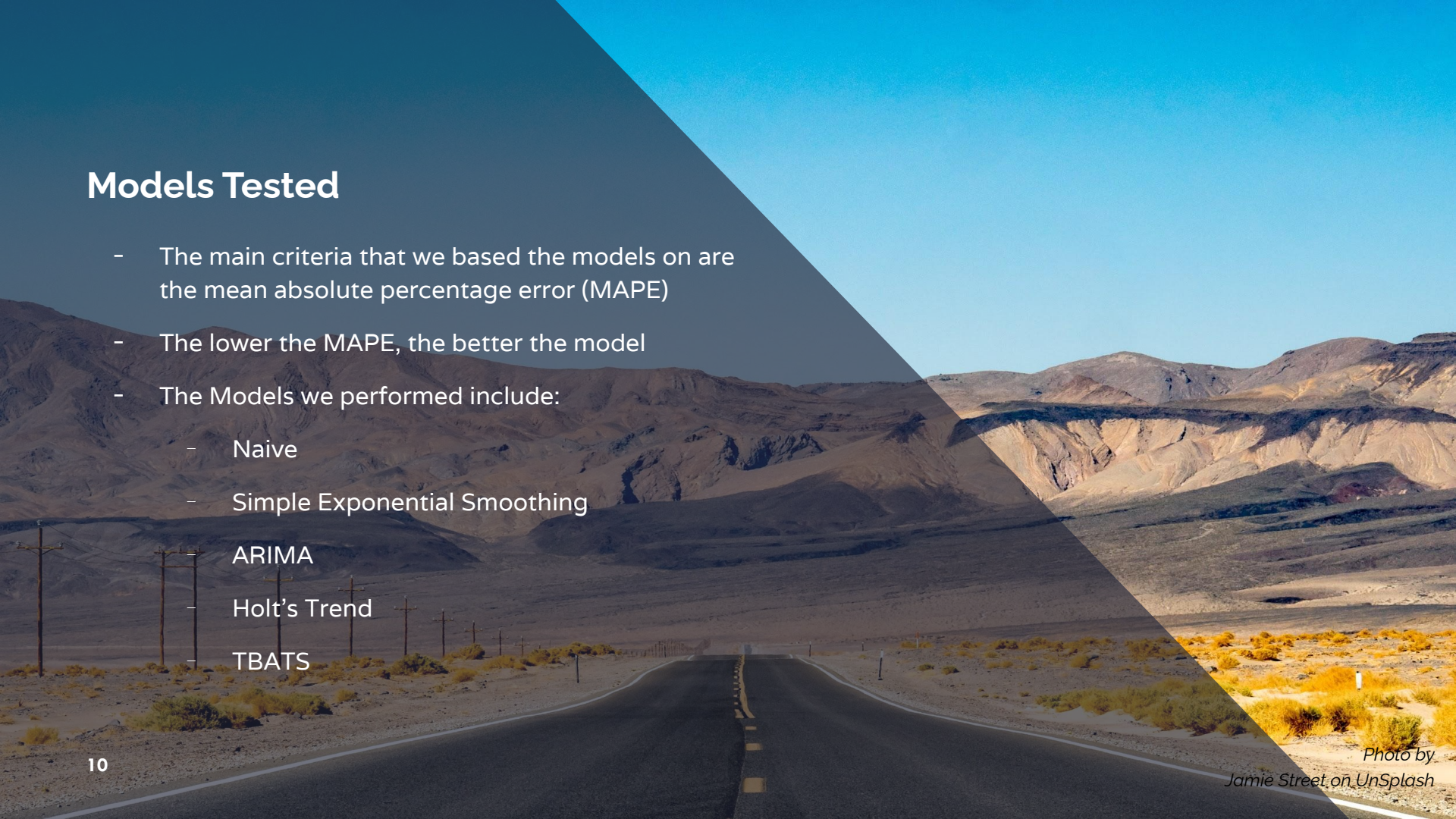


Photo by Jamie Street on UnSplash

MAPE Results

Model	MAPE
Naive	5,629
SImple Exponential Smoothing	161
ARIMA	158
Holt's Trend	255
TBATS	162

Photo by Jamie Street on UnSplash

Chosen Model:
ARIMA (0,1,1)

Autoregressive Integrated Moving Average

- The ARIMA (0,1,1) model is the best because it has the lowest MAPE
- This is a type of smoothing model for stationary time series
- Parameters:
  - Autoregressive coefficient = 1
  - Differencing coefficient = 1,
  - Moving average coefficient = 1

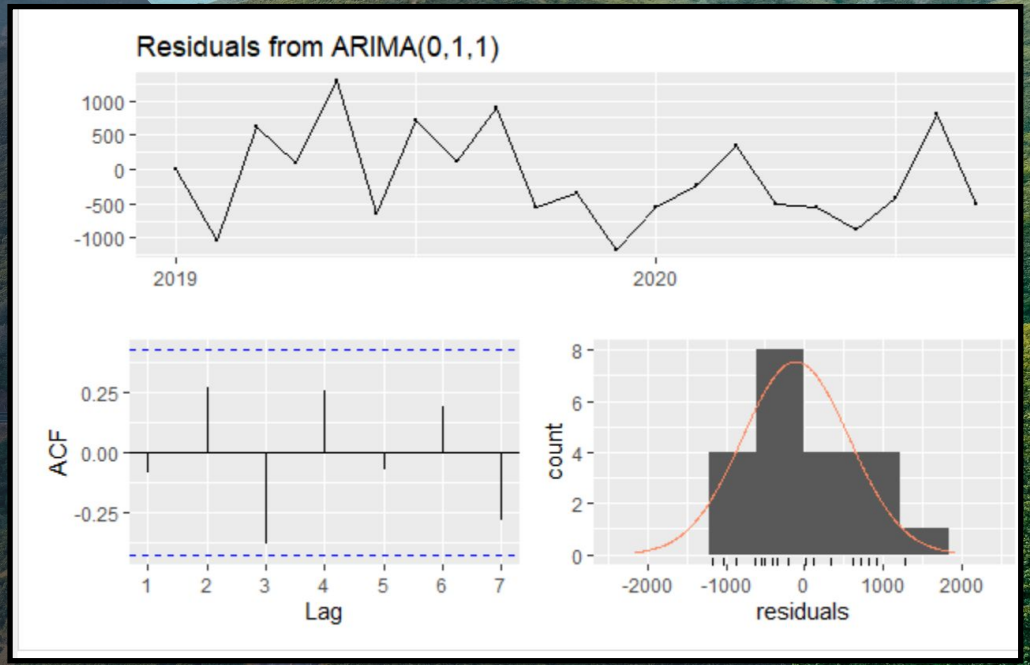
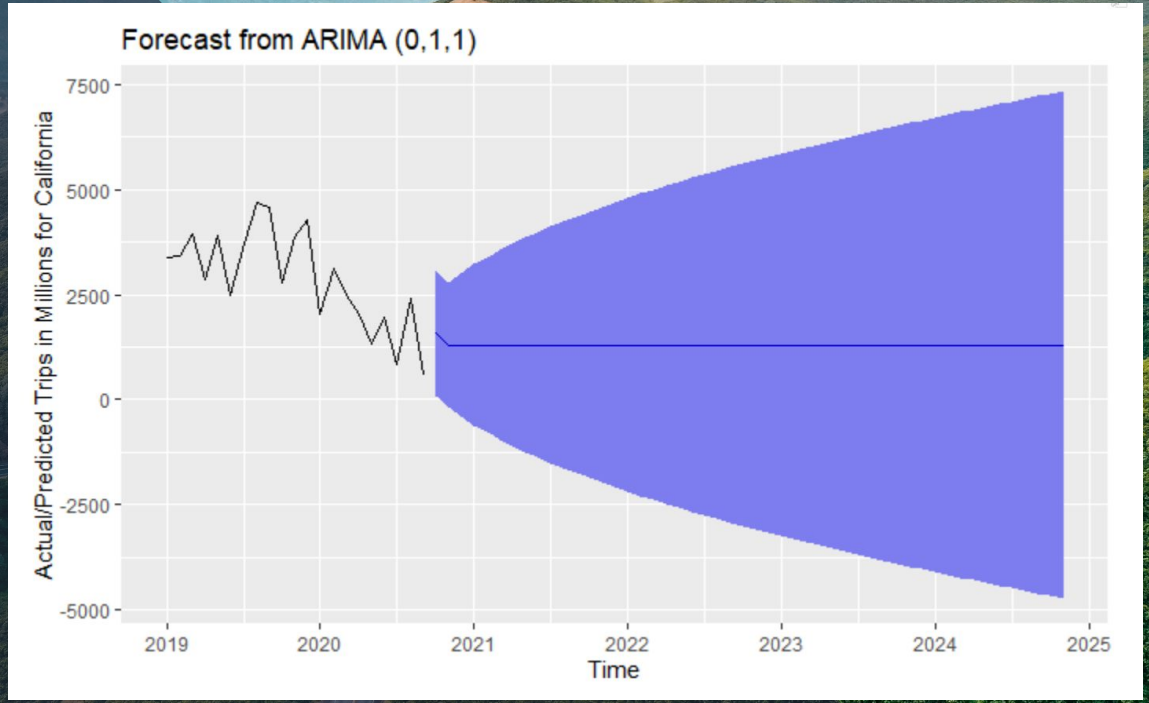


Photo by Drew Dau on UnSplash

Chosen Model: ARIMA (0,1,1)

Autoregressive Integrated Moving Average

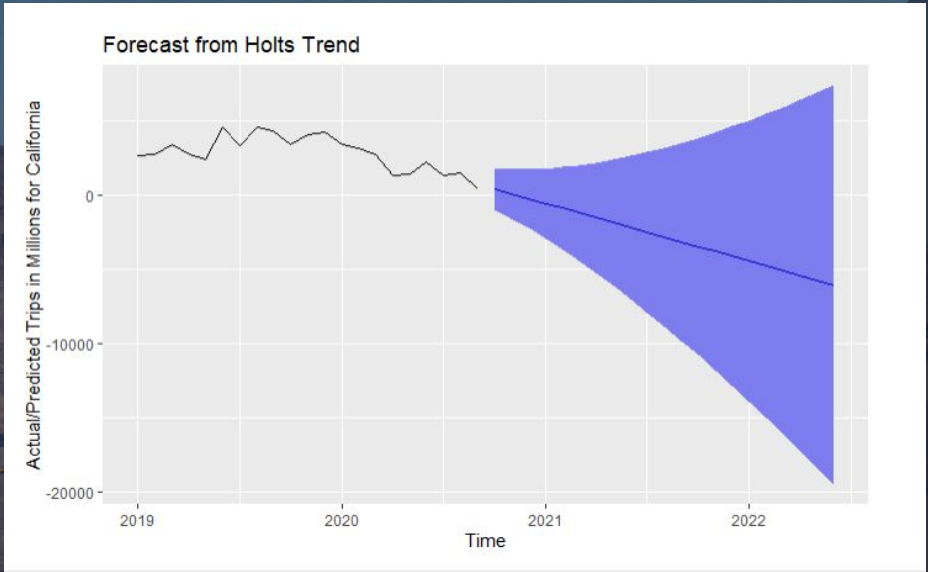
- The figure to the right shows us the predicted number of trips traveled into the future in millions
- It is estimated the number of trips per month will be around 1400 million but this will be hard to predict



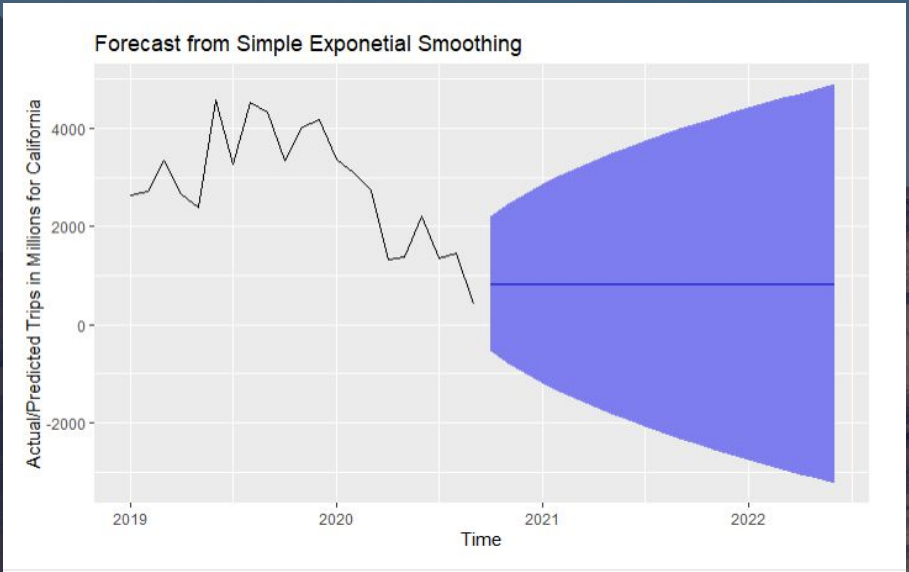
Black line is actual trip data. Blue line is predicted.
Purple shading is a 90% confidence interval

Photo by Drew Dau on UnSplash

Holt's Trend Results



Simple Exponential Smoothing Results



Black line is actual trip data. Blue line is predicted.
Purple shading is a 90% confidence interval

Photo by Jamie Street on UnSplash

3. SUMMARY

Assumption, Risk of Assumption and Conclusion

Photo by Nick on UnSplash

ASSUMPTION

We know the current trend of Covid-19 cases and deaths which will drive a lot of the travel decisions in the coming months. We saw a travel dip when Covid-19 first infiltrated California in March and April. If cases and deaths keep decreasing, we would expect travel to increase, but recently the Covid-19 numbers have continued rising.

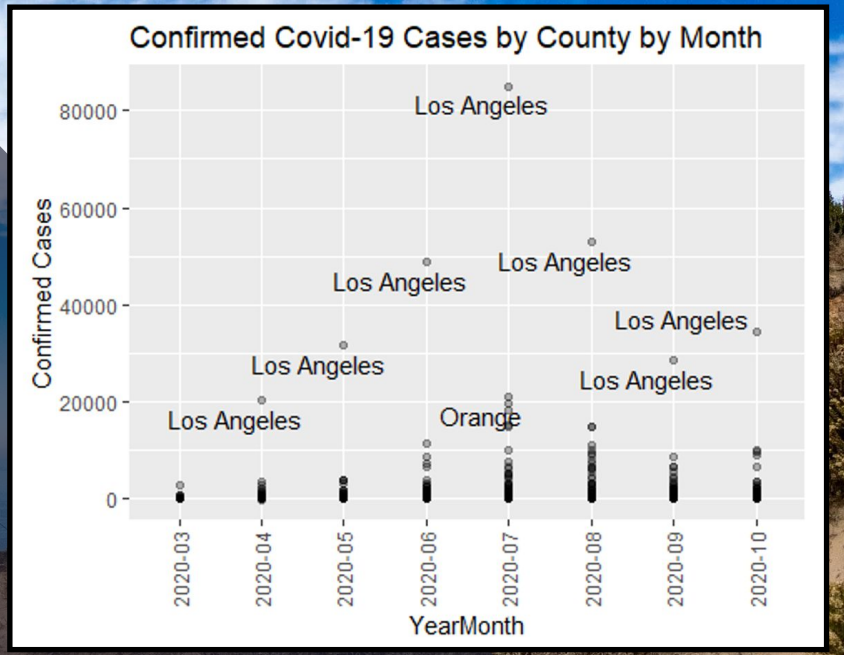


Photo by Constance Chen on UnSplash

RISKS OF ASSUMPTION

The assumption that Covid-19 cases will decrease and travel will increase is risky, especially with recent news of spikes. If flu season increases the amount of people sick, travel would decrease again. A working and distributed vaccine could decrease cases and deaths and increase the number of trips being taken. Overall, there are a lot of aspects that will affect travel in the future.

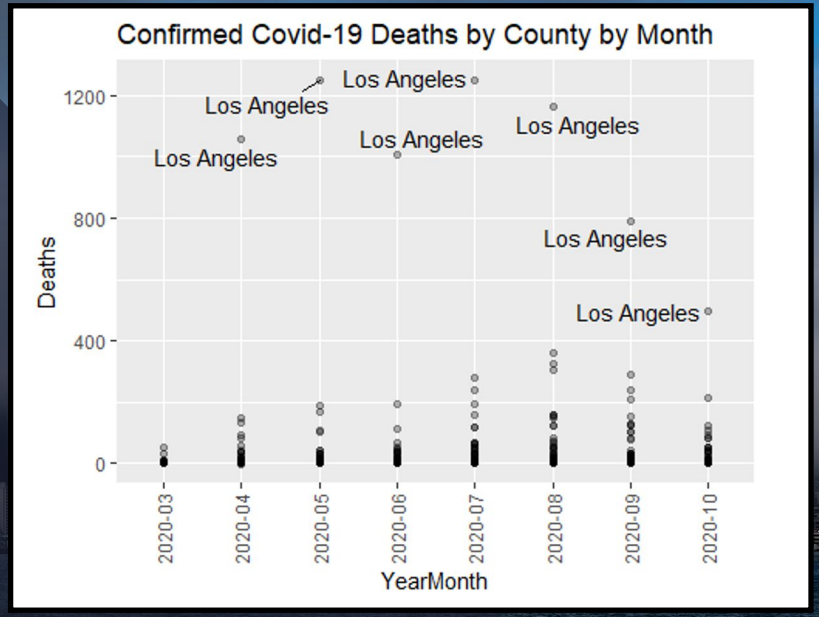


Photo by Casey Horner on UnSplash

CONCLUSION

- It is challenging to make predictions based solely on past trip numbers
- Using Covid-19 data helped with comparisons between trends observed and confirming expectations
- Weather, school breaks, income, and profession could all come into play when people are deciding to take trips or not
- Differentiating what kind of trips people are taking such as road trips, air travel, or train travel could offer valuable information for the transportation industry and travel agencies

Photo by Fabian Quintero on UnSplash

REFERENCES

BTS.gov. (2020, November 9). Trips By Distance. Retrieved from https://data.bts.gov/Research-and-Statistics/Trips-by-Distance/w96p-f2qv

California ALL. (2020, November 9). Tracking COVID-19 in California. CA.gov. Retrieved from https://covid19.ca.gov/state-dashboard/

California Open Data. (2020). COVID-19 Cases. Retrieved from https://data.ca.gov/dataset/covid-19-cases/resource/926fd08f-cc91-4828-af38-bd45de97f8c3

Data.gov. (2020, September 29). Trips By Distance. Bureau of Transportation Statistics. Retrieved from https://catalog.data.gov/dataset/trips-by-distance

Photo by Julian Myles on UnSplash

REFERENCES

Forecasting: Principles and Practice. (n.d.). Retrieved November 08, 2020, from https://otexts.com/fpp2/arima-r.html

Prabhakaran, S. (2020, September 17). ARIMA Model - Complete Guide to Time Series Forecasting in Python. Retrieved November 08, 2020, from https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/

Singh, D. (2019, July 12). Deepika Singh. Retrieved September 29, 2020, from https://www.pluralsight.com/guides/time-series-forecasting-using-r

Tourwriter. (2020). The travel industry is resilient: Historical events and how tourism has bounced back. Retrieved from https://www.tourwriter.com/travel-software-blog/covid-19-pt1/

Photo by Julian Myles on UnSplash

REFERENCES

Presentation template by SlidesCarnival

Photographs by Unsplash

Photo by Julian Myles on UnSplash