

*Kruze!*

# *Predicting Costs for Kruze! The Most Comfortable Airliner*

By Christopher Kuzemka

*Kruze!*



# Takeoff!

Kruze!



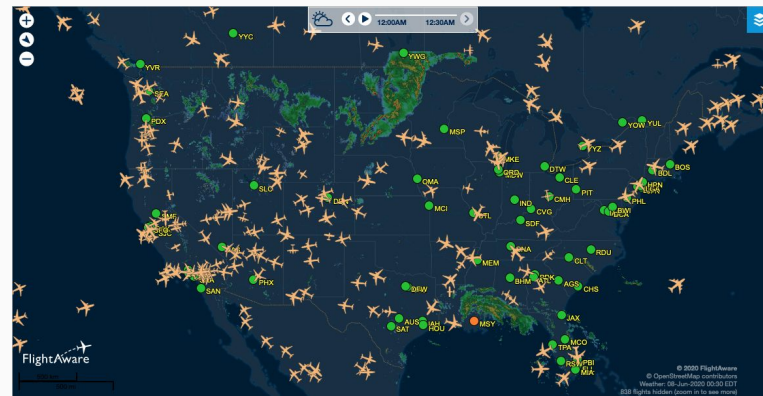
Consider the case example where a startup airliner, known as "Kruze", wants to establish itself as a top competitor against existing airliners today. A part of this startup process focuses on understanding the costs that will come into play when managing flights. Our job as data scientists today is to help Kruze determine the minimum threshold cost the airliner must charge their passengers in order to break even with a profit. To do this, we are going to analyze seven different locations where Kruze would like to establish themselves and examine existing flight route data as well as existing flight ticket prices (as a prediction) to help us create a supervised learning model.

As we are working with what is considered to be a continuous variable, we will analyze common price trends utilizing a supervised regression model, such as Linear Regression, KNNRegression, Decision Tree Regression, and Bagging Regression. We will ultimately be using the Mean Absolute Error against our predictions to help us gauge how well our selected model predicts the price and discuss what issues may be observed from the limitations of this study.

## How Can We Help?

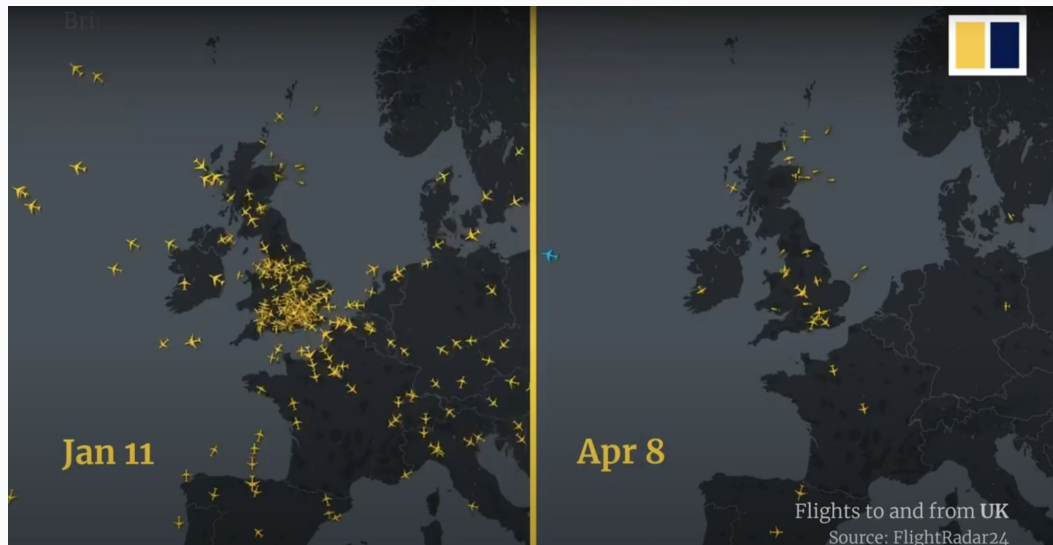
### Key Airports

- New York John F. Kennedy Airport (KJFK)
- Chicago O'Hare International Airport (KORD)
- Los Angeles International Airport (KLAX)
- Houston George Bush Intercontinental Airport (KIAH)
- Miami International Airport (KMIA)
- Hartsfield-Jackson Atlanta International Airport (KATL)
- Portland International Airport (KPDX)



# Complications and Limitations

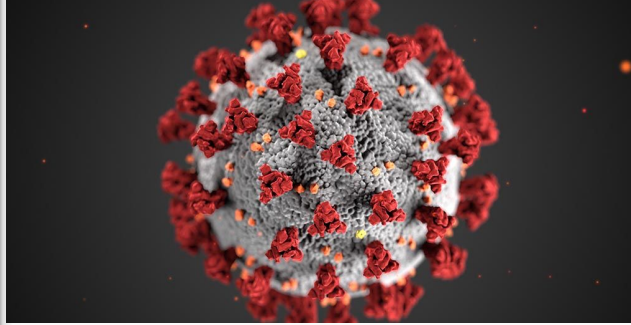
- API restrictions
  - Lack of seasonality
  - Lack of price gridsearching
  - No engine data
  - Dirty data nuances
- API Pricing (\$\$\$)
- Lack of relevant data
  - Jet fuel pricing
  - Maintenance costs
  - Crew salaries
  - Plane housing
- **COVID-19 Pandemic**



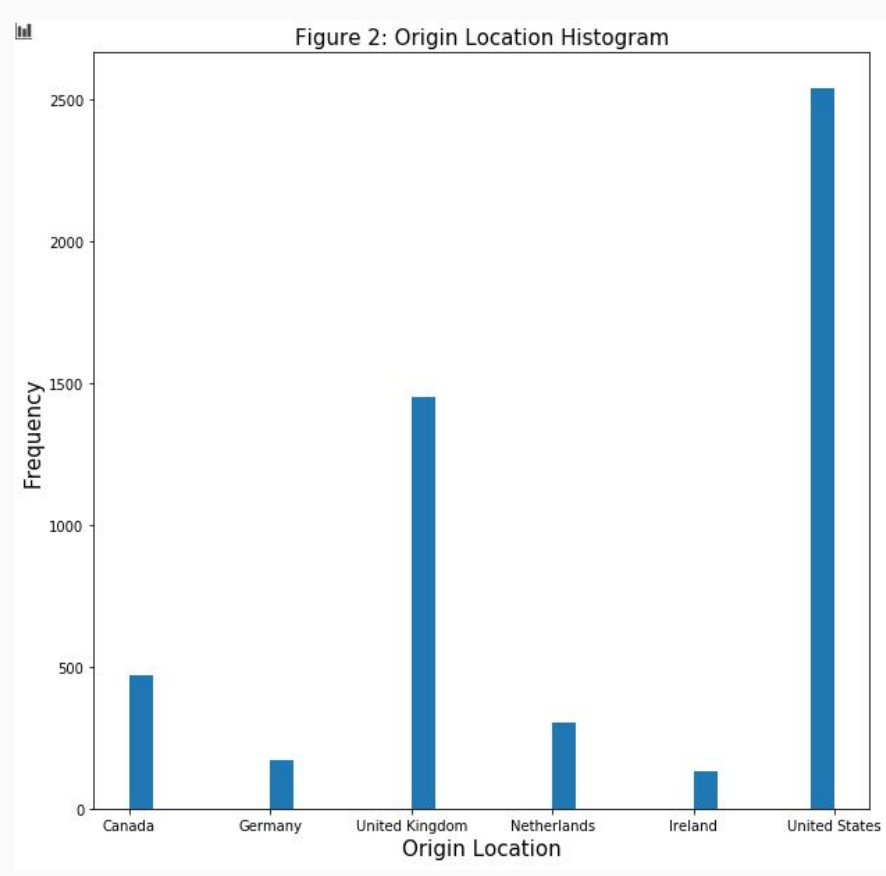
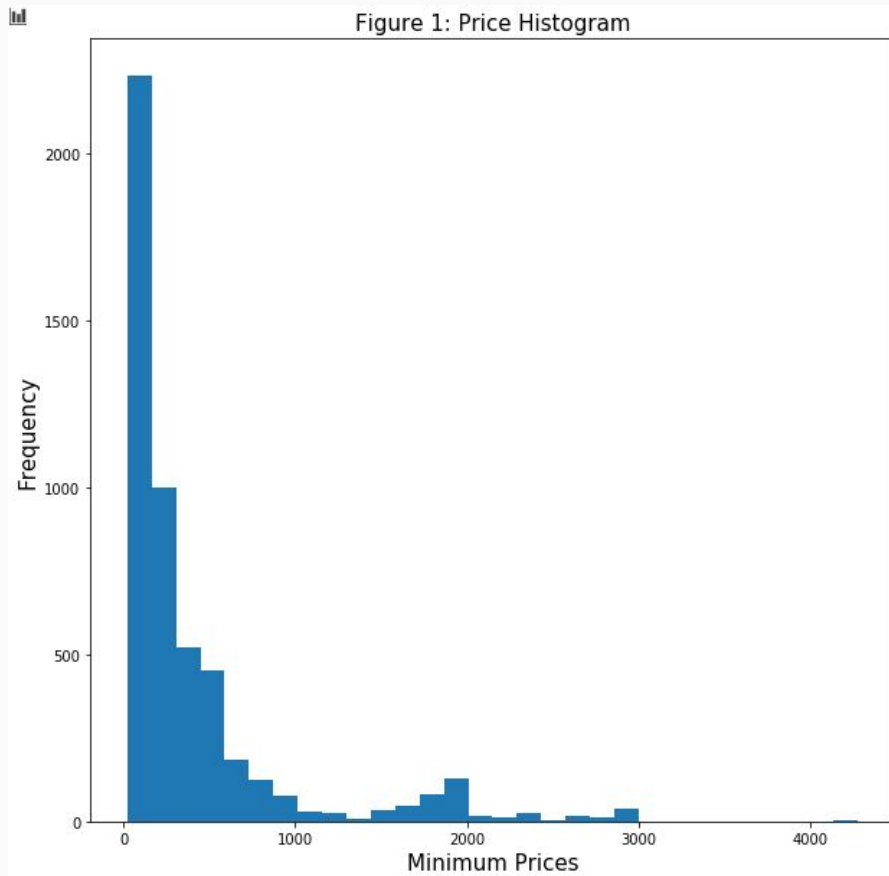
# The Pandemic

Largest and most important inhibitor on gathering data came from the Covid-19 pandemic. Air traffic around the world was slowed-down or brought to a halt.

Dramatic lack of flights for even popular routes. Stunning numbers!



Date	Total Traveler Throughput	
	Total Trave	(1 Year Ago - Same Weekday)
5/26/20	264,843	2,453,649
5/25/20	340,769	2,512,237
5/24/20	267,451	2,070,716
5/23/20	253,190	2,124,825
5/22/20	348,673	2,792,670
5/21/20	318,449	2,673,635
5/20/20	230,367	2,472,123
5/19/20	190,477	2,312,727
5/18/20	244,176	2,615,691
5/17/20	253,807	2,620,276
5/16/20	193,340	2,091,116
5/15/20	250,467	2,664,549
5/14/20	234,928	2,611,324
5/13/20	176,667	2,343,675
5/12/20	163,205	2,191,387
5/11/20	215,645	2,512,315
5/10/20	200,815	2,419,114
5/9/20	169,580	1,985,942
5/8/20	215,444	2,602,631
5/7/20	190,863	2,555,342
5/6/20	140,409	2,270,662
5/5/20	130,601	2,106,597
5/4/20	163,692	2,470,969
5/3/20	170,254	2,512,598
5/2/20	134,261	1,968,278
5/1/20	171,563	2,546,029
4/30/20	154,695	2,499,461
4/29/20	119,629	2,256,442
4/28/20	110,913	2,102,068
4/27/20	119,854	2,412,770
4/26/20	128,875	2,506,809
4/25/20	114,459	1,990,464
4/24/20	123,464	2,521,897
4/23/20	111,627	2,526,961
4/22/20	98,968	2,254,209
4/21/20	92,859	2,227,475
4/20/20	99,344	2,594,171
4/19/20	105,382	2,356,802
4/18/20	97,236	1,988,205
4/17/20	106,385	2,457,133
4/16/20	95,085	2,616,158



# Model Training

## **Baseline Model**

*Mean Absolute Error of \$368.*

## **KNN Regression**

*Mean Absolute Error of \$215.*

## **Linear Regression**

*Mean Absolute Error of \$123.*

## **Decision Tree Regressor**

*Mean Absolute Error of \$0.104*

*Kruze!*



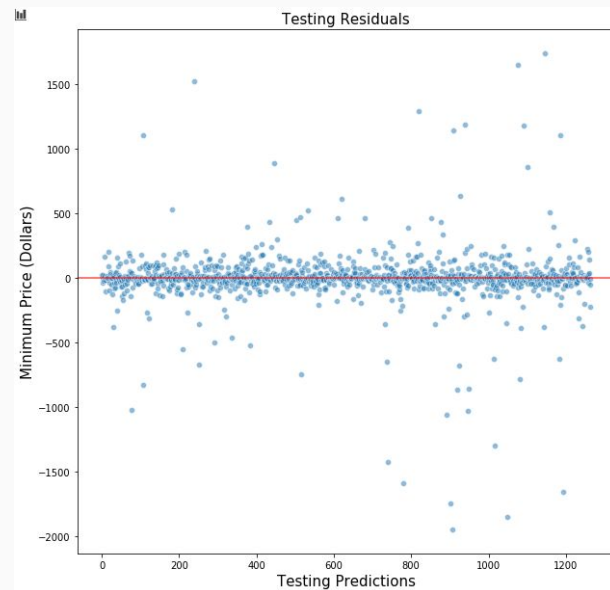
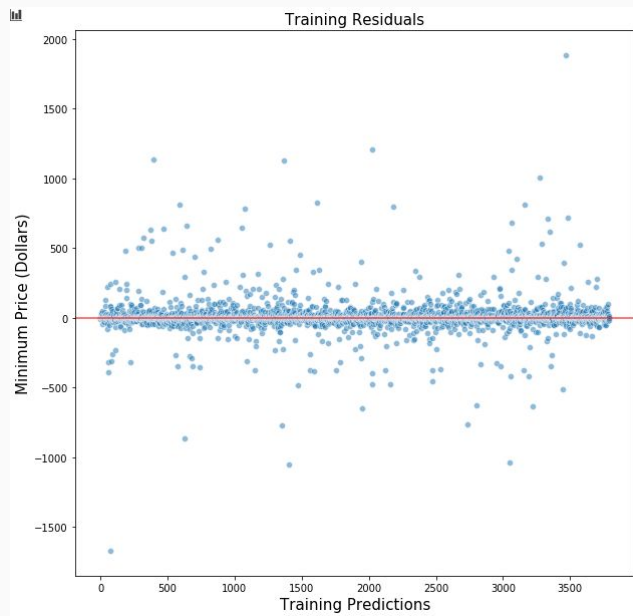
*Kruze!*

# Best Performing Model

**Bagging Regression**

*Training Mean Absolute  
Error of \$35.*

*Testing Mean Absolute  
Error of \$85.*





# Future Work

- More data collection
- More time dedication
- Incorporate Brayton Thermal Engine analysis
- Consider coordinated planning on flight searches and quote searches
- Jet fuel pricing
- Different models
- More features
- Pivot to application

