* **Slide 1: Cover Slide**
* **Slide 2: Context** 
  + Buyer’s Choice (BC) is a small used car company based in CT lead by CEO Dapper Dan. BC's usual process for deciding to buy vehicles from its warehouse is to determine what the likely price would be if the vehicle was sold in the store (BC Price).
  + If the BC predicted price is more than $2k than the warehouse asking price, then BC buys the vehicle. They expect to run a profit at least 80 percent of the time.
  + However with the advent of Covid, people just are not coming into the store. BC has therefore switched to selling their cars online. These prices often differ from the prices they would normally get at the store. BC has therefore created an algorithm to predict the value of a vehicle.
  + The company has recently been given the option to buy 25 vehicles. BC wants to know which of these 25 vehicles they should purchase.
* **Slide 3: Problem** 
  + Should used car dealership Buyer’s Choice buy any of the gently used vehicles being offered?
* **Slide 4: Aims and Objective** 
  + To answer this question we need to predict the value of each used car. If the predicted price is more than $2k than the asking price, BC should buy the vehicle.
  + Our analysis also explores what factors affect online used vehicle prices?
* **Slide 5: Intro to Dataset** 
  + Sources of Data
    - Craiglist
    - Back4App
    - MSRP
    - State Gas
    - Household Income
  + Using these sources, I created a dataset
    - Comprised of 19 variables [141 when some categorical variables were turned into dummies)
    - Totaling 114810 number of vehicles
* **Slide 6-8: Data Cleaning Rule**
  + Data is incomplete
    - Price, make, model, model year, mileage or condition are unlisted
  + The study is not looking at non super luxury vehicles passenger vehicles
    - MSRP < $500K
    - Price < $100K
    - Number of Cylinders <12
    - Vehicle is not a motorbike or bus
  + This study is looking only at passenger vehicles in good or better condition
    - Condition is not listed as poor or worse
    - Mileage < 200K miles
    - Average Mileage< 36K miles per year
    - Age is between 0 to 10 years old
* Slide 8b Other Variables
* **Slide 9-10: Exploratory Data Analysis (Descriptive Visuals)**
  + Age vs Mileage
  + **What Does Price Look Like?**
    - Age, Mileage
    - Trim (cylinders,drive, type)
    - Category Price
* **Slide 11: Pre-Processing** 
  + Standard Scaler
  + Creating Dummies
* **Slide 12: Modeling (Choosing the Correct Algorithm)** 
  + Tested 10 regression models
    - Linear Regression
    - Linear Log Regression
    - Ridge Regression
    - Elasticnet Regression
    - Decision Tree Regression
    - Extra Tree Regression
    - Gamma Regression
    - Ridge with PCA
    - Linear with PCA
    - elasticnet with PCA
* **Slide 13: Evaluation Metrics** 
  + Tested with two parameters
    - Median absolute error (how accurate it is)
    - Mean percentage error (how much it overestimates vs. underestimates)
      * Model is risk acerse
  + Error highly correlated with price; biased
* **Slide 14: MPE Comparison**
* **Slide 15: MAE Comparison**
* **Slide 16: Winner: Random Forest (Algorithm)**
* **Slide 17: Top Factors**
* **Slide 18: Conclusion**
  + **Prediction: Should Buy 9 of the vehicles**
  + **Confidence: +/- $2k**
* **Slide 19: Assumptions, Limitations and Disclaimers**
* **Slide 20: Next Steps / Direction**
* **Slide 21: Credits**

**\*Random forest Test Accuracy**

**\*Apply model to say on validation**