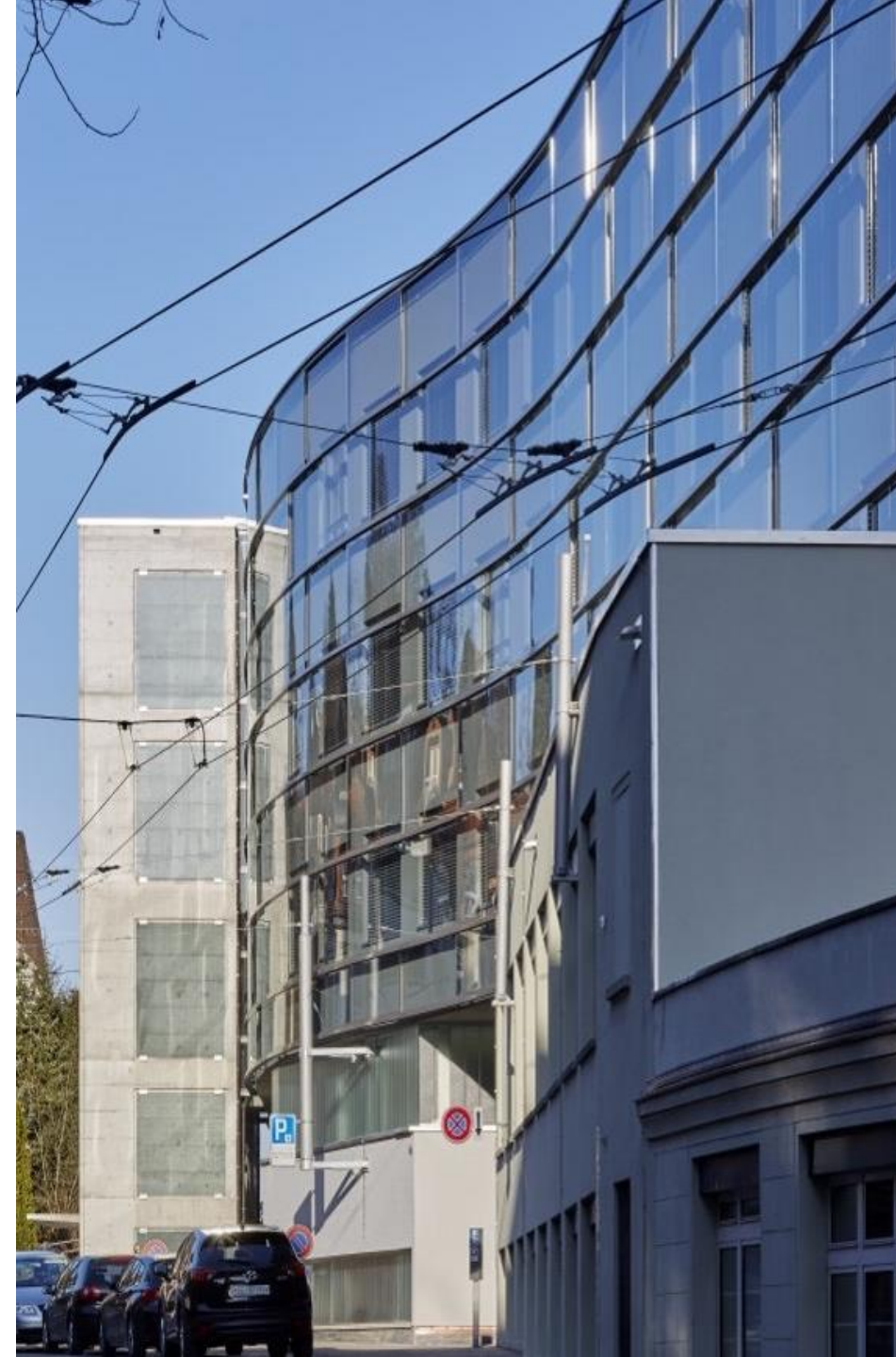


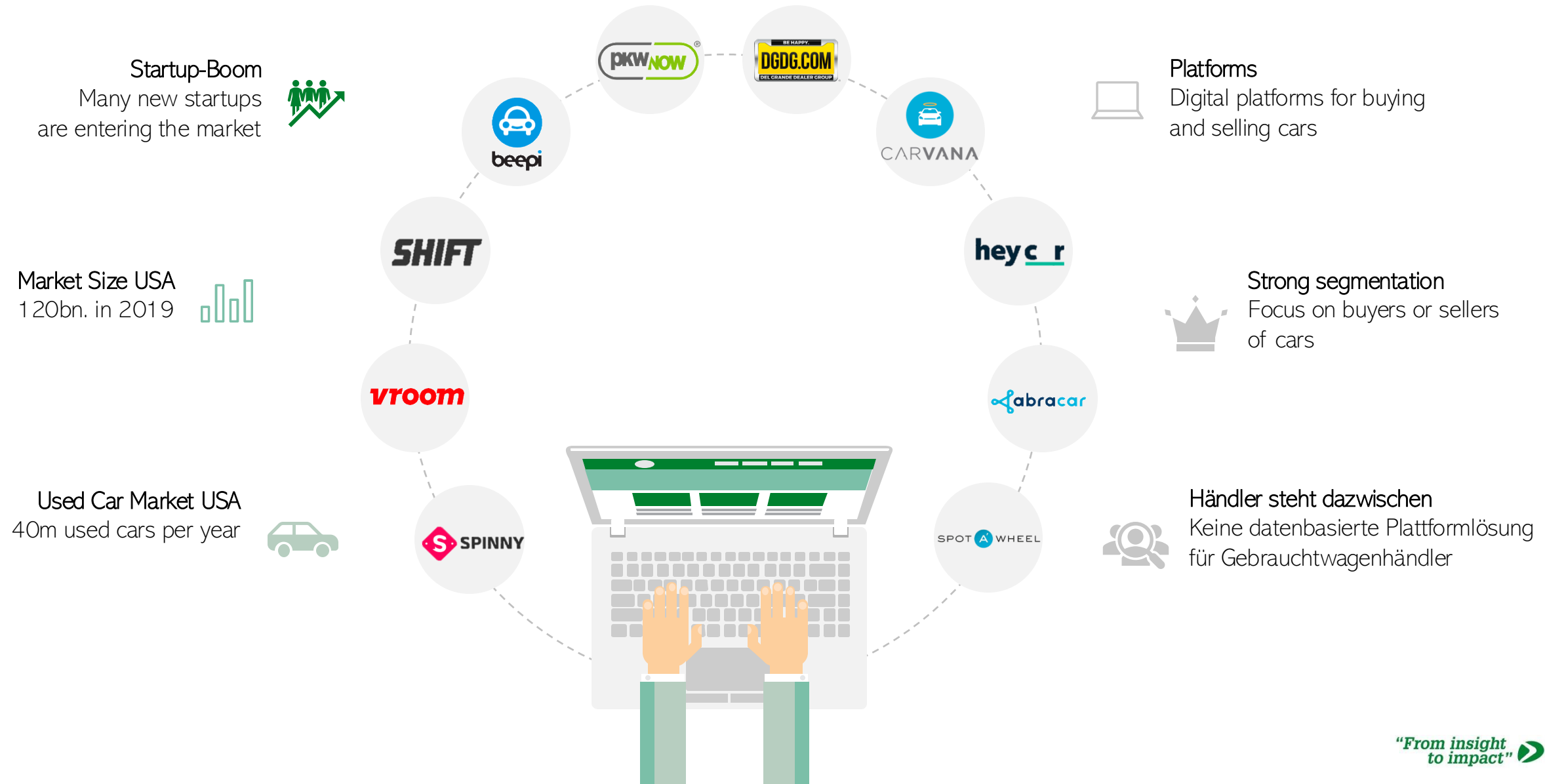
# Assignment 2

Prof. Dr. Ivo Blohm





# Digital Transformation of the Used Car Market

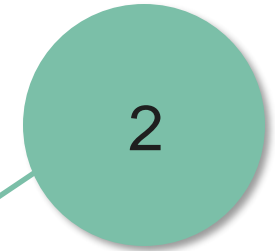
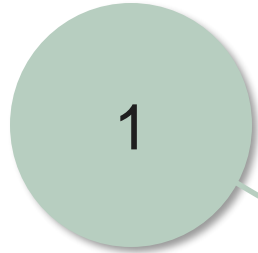




# The promises of Price Predictions

**Purchasing Cars**  
Buy at a fair prices

**Selling Cars**  
Don't sell your car under price



Prediction



# The Data Set

- The Data Set has been crawled from Craigslist
- Craigslist is an US-american classified advertisement websites with sections devoted to jobs, cars, private sales, etc.
- Data set contains **all historic car sales listings**
  - >500k Listings
  - 1.45 GB data
  - Regularly updated
  - 26 „typical variables“ to describe a car (manufacturer, status, km, etc.)



We use data on  
~ 31k listings



# We did some Pre-Processing

## Handling Outliers

- Deletion of instances with more than 5 instances
- Deletion of cars with \$0
- Deleting outliers < \$650 und > \$150k

## Deletion of less important / faulty variables

- Region
- id
- url
- image\_url
- region\_url
- vin
- description
- county
- lat
- long

## Imputing NAs

Deleting instances with missing values for year, manufacturer, model, fuel, title\_status & transmission (very few instances)

Imputing missing values for features cylinders, condition, drive, size, type, paint\_color (takes into account correlation between variables)



Random Sampling of Instances





# Ihre Aufgabe

- Craigslist hat sich entschieden in das Gebrauchtwagengeschäft einzusteigen
- Es soll ein Modell entwickelt werden, mit denen Gebrauchtwagenpreise vorhergesagt werden können.
- Sie wurden engagiert dieses Model zu entwickeln.



## Assignment 3 - Ihre Aufgabe

- Trainieren Sie drei verschiedene Modelle mit unterschiedlichen Algorithmen zur Vorhersage von Gebrauchtwagenpreisen!
- Wenden Sie bei einem Modell Grid Search und Cross Validation an, um die Hyperparameter zu tunen!
- Entwickeln Sie eine Visualisierung, mit der Sie die Performance der drei Modelle vergleichen können!
- Überlegen Sie sich eine konkrete Empfehlung, wie Sie die Vorhersagequalität versuchen würden weiter zu steigern.
- Fassen Sie Ihre Arbeit in 2-3 Powerpoint-Folien zusammen. Gehen Sie dabei auf Ihre Vorgehensweise und Ihre Ergebnisse ein.



# Teilnahme an Inclass Kaggle Competition

- Wie gut sind die eigenen Vorhersagen im Vergleich zu denen der anderen Teilnehmer?
- Let's find out!



HAPPY  
CODING