Sample structure:

* Intro – 10 - 2 – A
  + Intro – map overview state -> NYC
  + Business case
* Data – 25 – 5 - A
  + (Ratings /) Spatial data
  + Web scraping
  + Dem <- hidden correlativity
  + Airbnb
  + Subway
  + Overview
* Methods – 35
  + Balancing <- model always predicts As
  + Resampling -> different outcomes & minima’s -> over/under-bagging
  + Model selection procedure
    - Bagging sample
    - Cross validation
    - Variable selection & tuning
  + Problem slide: best subset selection -> would take a year
  + Tried
    - Out of bag errors -> testing data unbalanced, misleading results
    - Function: Avoided for loops with apply,
    - Best subset -replaced by-> forward stepwise selection -> less computing time
  + General problem: computational intensity
* Models & Results – 30 – K
  + Try to present as overviews with comparisons
  + Write the model-makers what you need from them exactly
* Conclusion – 5 – K
  + Independent variables
  + -> better opportunities for agencies

Ideas:/comments

* Lots of graphs
* Code snippets?
* Better many slides with light info density then the other way around
* But: comparisons of different stuff in one slide nice
* Include an overview of all the data sources we used <- to illustrate how many different approaches we had to use
* Should minimize swithes (not 9)
* Be prepared for questions about business case

Script Andre:

1. State map
   1. Welcome
   2. Short description
2. City map
   1. Why:
   2. 48 million people fall sick, 128’000 are hospitalized and 3’000 <- estimated
   3. Food Safety and Inspection Services
      1. Over 9k employees
      2. Budget: over billion $
      3. NY field: 13-15 million $
   4. FDA
      1. Over 16k
      2. Budget: Over 5 billion $