**315 Project Timeline**

* **Dataset:** <https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data#AB_NYC_2019.csv>
* **Variables:**
  + info about airbnb: `id`(nominal), `name`(nominal)
  + Host Info: `host\_id`(nominal), `host\_name`(nominal), `calculated\_host\_list`(numerical, amount of listing per host)
  + Geographic Info: `neighbourhood`(categorical), `latitude`(numerical), `longitude`(numerical)
  + Basic Airbnb Info: `room\_type`(categorical), `price`(numerical), `minimum\_nights`(numerical, amount of nights minimum to book), `availability\_365`(numerical, number of days when listing is available for booking)
  + Popularity Metric: `number\_of\_reviews`(numerical), `last\_review`(date-time-year, time of latest review), `reviews\_per\_month`(numerical, number of reviews per month)
* **Questions we like to explore?**
  + Which area and room type have more popular listings?
  + What variables affect prices
* **Part1: EDA [plot + interpretation] due Friday Nov 29**
  + Histogram of Price [Lucy]
    - Stacked Histogram/ Side-by-side Histogram
    - Color by neighborhood
  + Correlation Network [Lynn]
    - Igraph package
    - Refer to HW9 1a
    - See which variables are connected to price, try different significant level
  + Pairwise Correlation Graph [Lynn]
  + Pairwise Scatter Plot [ Lynn ]
  + Correlation Heatmap: [Lucy]
    - Refer to HW9 1a
  + Parallel Coordinate Graph & Radar Graph [Lucy]
    - (on Numerical Variables)
    - Color by neighbourhood
    - Refer to HW8 1a
  + Word cloud of hosting name [Chen Xi]
    - (determine most popular words for house listing)
    - Facet by neighborhood\_group
  + Map of NYC [Serena]
    - Each neighborhood\_group is colored by mean price range in that area
    - Add dot to each neighborhood\_group, size by number of listings in this neighborhood
    - Refer to HW10
  + Dendrogram on all numerical variables: [Serena]
    - Refer to HW8 3b
    - See which variables are more related
* **Explore how each variable is related to prices (decide after EDA) Due Sunday Dec1**
  + Zoom-in Scatter Plot: [Chen Xi]
    - Relationship between two variables (color/ shape), then zooming in
    - Refer to HW8 7
  + Linear Regression: [Chen Xi]
    - Geom\_smooth
    - Refer to HW6 problem 4
* **Make Poster + Submit Due Monday (Remote)**