# Work allocation

**Person 1 (Pitshou): Introduction & Business Question (150-200 words)**

* Write Introduction (context of Airbnb, why NYC dataset matters)
* Business question
* Provide rationale (why this matters for Airbnb)
* Explain the expected business impact

**Person 2 (Gayathri): Methodology (350-400 words)**

* Handle data preparation & cleaning (missing values, outliers, demand proxy)
* Explain feature engineering (portfolio size, demand variable, neighbourhoods)
* Describe regression approach (for pricing) in plain business terms
* Describe clustering approach (for segmentation) in plain business terms
* Include code in Appendix

**Person 3 (Theodoros): Results & Visualisations (Regression) (250 words)**

* Present EDA findings (distributions, boxplots, scatterplots)
* Show regression results → keep business-friendly!
* Use visuals with short business interpretation takeaways and captions (maybe 2 visuals?)

**Person 4 (Nilus): Results & Visualisations (Clustering + Conclusion) (250 words)**

* Present clustering results → describe host/customer segments in plain language.
* Add 1–2 visualisations (like scatter plot showing clusters?)
* Write concise insights for discussion & Conclusion (business insights, recommendations).

# Business Topic

**Business Topic:** How does the scale of a host’s portfolio (number of listings) influence Airbnb pricing and demand trends across neighbourhoods in NYC?

**Reasoning:**

* Airbnb has both casual hosts (1 listing) and professional hosts (multiple listings)
  + These groups often differ in pricing strategies, demand levels, and neighbourhood presence
  + Makes the topic is well suited to the NYC dataset (contains host listing counts, prices, demand indicators)
* Analysing portfolio scale shows whether professional hosts have an advantage in pricing or demand
* Insights are directly relevant for Airbnb’s strategic planning and host policies