# CapstoneProjectSubmission

Instructions:

1. Pleasefillinalltherequiredinformation.
2. Avoidgrammaticalerrors.

TeamMember’sName,EmailandContribution:

# Vineeta Singh([singhvineeta0118@gmail.com](mailto:singhvineeta0118@gmail.com)):

* 1. DescriptiveAnalysis
     1. Dataframedescription
     2. Dataframeshape

Data cleaning and Analysis:

* + 1. Dropunwantedcolumns[‘ID’,’lastreview’]
    2. Box Plot of price
  1. DataWrangling
     1. Relationbetween‘Room\_type’and‘minimum\_nights’
     2. Relationbetween‘Room\_type’and‘number of reviews’
     3. Relationbetween‘Neighbourhood group’and‘number of reviews’

Data Visualisation

* + 1. RelationshipbetweenRoom\_type’and‘minimum\_nights’
    2. RelationshipbetweenRoom\_type’and‘Number of reviews’
    3. RelationshipbetweenNeighbourhood\_group’and‘number of reviews’

# Tushar Gupta ([gupta.tushar1992@gmail.com](mailto:gupta.tushar1992@gmail.com)):

Descriptive Analysis

* + 1. Dataframedescription
    2. Dataframeshape

Data cleaning and Analysis

* + 1. RemovingNaNvaluesfromcolumnslikePrice and Host Name
    2. Correlation Plot of all variables
    3. Box plot between price and neighborhood group

Data Wrangling

* + 1. Relationbetween‘Neighbourhood\_group’and‘most number of host’
    2. Calculationofaverageprice per neighborhood group
    3. Maximumbenefittedhost

Data Visualisations

* + 1. Relationshipbetween‘Neighbourhood\_group’and‘number of host’
    2. RelationshipbetweenRoom\_type’and‘Neighbourhood\_group’

# Abhishek Mishra([abhishekmishra9559026@gmail.com](mailto:abhishekmishra9559026@gmail.com)):

Descriptive Analysis

Dataframedescription

Dataframeshape

Data cleaning and Analysis:

Handling null values of ‘reviews per month’

Find zero price columns and remove them

Removing null values from column name “name”

‘

Data Wrangling

Relationbetween‘Price’and‘reviews’

* + 1. Find top three busiest host
  1. DataVisualisation
     1. RelationshipbetweenRoom\_type’and‘review’ using scatter plot
     2. Plotting Graph of busiest host

# Mallesh ([kurvamallesh36@gmail.com](mailto:kurvamallesh36@gmail.com)):

DescriptiveAnalysis

Dataframedescription

Dataframeshape

Data Wrangling

* + 1. Relationship between top three host and each neighbourhoodgoup

Room types preferred in Manahattan and Brooklyn

Data Visualisation

* + 1. Relationship between top busiest hosts in each neighbourhood group

Geographical graph in neighbourhood

|  |
| --- |
| Arunesh Mishra([Arunesh12mishra@gmail.com](mailto:Arunesh12mishra@gmail.com)): DescriptiveAnalysis   * + 1. Dataframedescription     2. Dataframeshape   1. DataWrangling   Room types preferred in Manahattan and Brooklyn   * 1. DataVisualisation   Geographical graph ofneighbourhood   * + 1. Relationship between busiest host in each neighbourhood group |
| GithubRepositorylink:  **MeghaVashist**- <https://github.com/megha-mv/AIRBNB-PROJECT-1-CAPSTON.git>  **Rohit bhadauriya**-<https://github.com/Creatrohit9/AIRBNB-PROJECT-1CAPSTON.git>  **SiddhantBashi**-https://github.com/Siddhant469/AIRBNB\_CAPSTONE01.git |

Please write a short summary of your Capstone project and its components. Describe the problemstatement,yourapproachesand yourconclusions. (200-400 words)

Airbnb (AirBed&Breakfast,where Air was nothing but mattresses) comes with a small idea ofearning rent by Brian Chesky and Joe Gebbia in 2007 turned hospitality and travel industry on itshead and gained massive recognition now. In this Capstone Project, we were given dataset of 49000observationsdistributedin16columns.Columnsareoftwotypescategoricalandnumericaltype.

Initially we started with analyzing the data. And after some initial analyzing ,it was gathered thatsome columns are not useful and then we drop them.And some columns have NaN(null) values whichalso need removal.Thiswholeprocesswastermed asDataCleaningand Analyses.

Next moving forward ,started with Descriptive analysis over the raw data provided using box plot, correlation and other descriptive methods. The Project isnowfurther brokenintomajorlyfourpartsaccordingtothequestionwe need toanswer.

At first, hosts and room\_types were analysed.From that it was concluded that ‘Entire home/apt’ weremostpreferredtypeofroom.Onfurtheranalysis,usingdatavisualizationtechniquesaclearrelationship between room\_type and minimum number nights spent was shown,which concluded thatiftherooms typeis‘Entirehome/apt’ thencustomersspentmorenumberofnightsthere.

Nextafteranalyzingtheneighbourhoodgroupwithrespecttoaveragepriceateachlocation,Manhattan was concluded as costliest state among others and possible reason could be the financialstateof entirecountry. People are not money conscious when spending money mainly in Manhattan and Brookyln as they want more privacy

And then analyzing was done based on ‘number of reviews and host listing’ which provided host were giving more reviews in most listed neighbourhood

Also we looked for the top three busiest host in US and also in top three busiest in each neighbourhood group

After exploring more about neighbourhood,we can conclude that shared room type were less likelypreferredbyUScustomersofAirbnb.AndBedford-stuyvesentandWilliamsburgweremostpopulatedfromManhattan state.

WecanconcludethatmostoftheAirbnbhostsarepresentinManhattanandBrooklyn.Aftercorrelation between price and neighbourhood,