Presentation Content: Walking in Paris Team

```
Objective:
Identify the attributes that make Airbnb “Successful”
Find the best place to stay in Hawaii

```

```
Hypothesis:
Ho:
The number of properties rented out by a host has no impact on rating.

Null Hypothesis:
H1:
None of the factors we are analyzing (neighborhood, number of listings per host, room type, price, keywords in description, how long the host has been hosting, and certain amenities) have an impact on the overall rating.

```

```
Potential:
Use Google Places API to see if there is a relationship between "Major attraction", beach, and high bookings
```

Cleanning the data:

In [ ]:

```
List of the variables on the dataset. Comprise 106 columns, 23,746 rows.

The column named "id" is similar than the index.
    id
    listing_url
    scrape_id      --> Values not required
    last_scraped   --> Values not required
    name
    summary
    space
    description
    experiences_offered      --> Values not required
    neighborhood_overview
    notes
    transit
    access
    interaction
    house_rules
    thumbnail_url     -----> Column in blank
    medium_url        -----> Column in blank
    picture_url       -----> Values not required
    xl_picture_url     -----> Column in blank
    host_id
    host_url          --> Values not required
    host_name
    host_since
    host_location
    host_about
    host_response_time
    host_response_rate
    host_acceptance_rate      -----> Column in blank
    host_is_superhost
    host_thumbnail_url     --> Values not required
    host_picture_url       --> Values not required
    host_neighbourhood
    host_listings_count    --> Data duplicate on column host_total_listin
    host_total_listings_count
    host_verifications        -----> Column in blank
    host_has_profile_pic
    host_identity_verified
    street
    neighbourhood
    neighbourhood_cleansed
    neighbourhood_group_cleansed
    city
    state
    zipcode
    market
    smart_location     --> Values not required
    country_code       --> Values not required
    country
    latitude
    longitude
    is_location_exact    --> Values not required
    property_type
    room_type
```

```
    accommodates
    bathrooms
    bedrooms
    beds
    bed_type    --> Values not required
    amenities
    square_feet
    price
    weekly_price
    monthly_price
    security_deposit
    cleaning_fee
    guests_included
    extra_people
    minimum_nights
    maximum_nights
    minimum_minimum_nights
    maximum_minimum_nights
    minimum_maximum_nights
    maximum_maximum_nights
    minimum_nights_avg_ntm
    maximum_nights_avg_ntm
    calendar_updated    --> Values not required
    has_availability   --> Values not required
    availability_30    --> Values not required
    availability_60    --> Values not required
    availability_90    --> Values not required
    availability_365   --> Values not required
    calendar_last_scraped   --> Values not required
    number_of_reviews
    number_of_reviews_ltm
    first_review
    last_review
    review_scores_rating
    review_scores_accuracy
    review_scores_cleanliness
    review_scores_checkin
    review_scores_communication
    review_scores_location
    review_scores_value
    requires_license   --> Values not required
    license            --> Values not required
    jurisdiction_names  --> Values not required
    instant_bookable   --> Values not required
    is_business_travel_ready   --> Values not required
    cancellation_policy
    require_guest_profile_picture  --> Values not required
    require_guest_phone_verification  --> Values not required
    calculated_host_listings_count
    calculated_host_listings_count_entire_homes
    calculated_host_listings_count_private_rooms
    calculated_host_listings_count_shared_rooms
    reviews_per_month
```

```
114
```

In [ ]:

```
Cleaning raw data (airbnb) to delete the column with not values, columns
1.- The following columns has been drop from the analysis:
    "thumbnail_url",
    "picture_url",
    "medium_url",
    "xl_picture_url",
    "host_url",
    "host_thumbnail_url",
    "host_picture_url",
    "scrape_id",
    "host_listings_count", # There is a host_total_listings_count that i
    "host_acceptance_rate", # NaN
    "calendar_last_scraped",
    "bed_type",
    "last_scraped",
    "calendar_updated",
    "has_availability",
    "availability_30",
    "availability_60",
    "availability_90",
    "availability_365",
    "license"

2.- Drop the row on the columns "number of reviews" equal to zero or bla
    'review_scores_rating',
    'review_scores_accuracy',
    'review_scores_cleanliness',
    'review_scores_checkin',
    'review_scores_communication',
    'review_scores_location',
    'review_scores_value'

```

```
Analysis: Identify the variable that has more impact in the best value
and best price. Variable to be evaluate: neighborhood, type of place,
```

```
Meg:
Evaluate the relationship between "age of property" and
review_scores_rating
Evaluate the relationship between "price" and review_scores_rating
Evaluate the relationship between "how many properties someone has" and
review_scores_rating
Evaluate the relationship between "number of bedrooms" and
review_scores_rating
```

```
Armon:
Evaluate the relationship between "neighborhoods" and
review_scores_rating
```

```
Lance:
Evaluate the relationship between "Island" and review_scores_rating
Evaluate the relationship between "Type of place" and
review_scores_rating
```

```
Edith:
Build Logistic Regression
Consolidate Power Point
Consolidate Notebook
Write bullet point for cleansing data

```

```
Scott:
Audit the code and Technical review of the code
check the code is running and make any necessary adjustment
```

```
Edith and Meg prepare the final presentation with the input from the team
```

```
Team:
Write conclusions and recommendations
Each member write observationes from their own analysis
```

Type *Markdown* and LaTeX: α^2