Dhruv Shrivastava, Diego Victor Carlos Chavez, Gagan Pahuja, Padmapriya Panneerselvam, Ryan Egbert, Zainab Abdulla Aljaroudi

MGMT 585 – It Project Management

Airbnb Project Report – Project Mafia

**Contents**

[**Problem Statement:** 1](#_Toc84623689)

[**1. Application of Project Management to this Project** 1](#_Toc84623690)

[**2. Data Sources Summary** 3](#_Toc84623691)

[**3. Analysis** 3](#_Toc84623692)

[**4. Proposal for Further Research** 5](#_Toc84623693)

# **Problem Statement:** We did a current state, desired state and gap analysis for the problem.

**Current State:** Airbnb is a crowd-sourced online marketplace for lodging, homestays and vacation rentals which rose meteorotically ever since 2008. When the COVID virus spread over the world in March 2020, there was a reduction in bookings and Airbnb had to cost cut on multiple functions. Airbnb was faced with a situation where they have to operate in a new and changed world.

**Desired State:** The future state would be that Airbnb not only bounces back to the pre-covid demand but is able to make it’s bookings grow past it. This would happen by Airbnb making strategic decisions based on the new understanding of this changed situation. For such an understanding, Airbnb should have a strong hold of it’s understanding of what factors affect the bookings in the new world.

**Gap analysis:** The gap to be identified and filled here is the change in the driving factors for bookings pre covid and post covid. (Refer Figure1)

# **1. Application of Project Management to this Project**

**Our Approach**: We approached our project with Project Management methodologies: Waterfall and Agile. The Waterfall method is a simple sequential approach in which each project phase must be completed before beginning the next one, leading to the end deliverable. On the contrary, agile method comprises of short sprints that embrace flexibility and continuous improvement. In the waterfall methodology, we used MS project as opposed to Kanban boards in Agile.

**Waterfall (Plan-Based Approach)**: Initially, we created a Work breakdown structure and created summary tasks. After that, we estimated the task durations of each of the tasks and defined precedence relationships, followed by assigning lead/lag times between tasks. We then created the Gantt chart and PERT diagram and assigned resources to each of the task. After setting the baseline for our project, we entered the actuals and calculated the Earned Value Analysis.

**Earned Value Analysis:** Earned Value Analysis (EVA) is a Project Management technique for measuring project performance in an objective manner at any given point in time. EVA compares the budgeted cost of work scheduled (BCWS) with Actual cost of work performed (ACWP) and Budgeted cost of work performed (BCWP). We plotted the graph of how these three costs compared to each other throughout our project. BCWP = $11,845.00 and ACWP=$14102.50. See Appendix – Figure 3.

**Agile Approach**

**User stories:** Using INVEST methodology we create small independent and testable/actionable user stories with a description and an acceptance criteria for the expected outcome.

**MOSCOW Prioritization:** The MOSCOW methodology was used to decide which stories were relevant to the final outcome and could be done in the required time period and project scope.

**Planning Poker:** The team played planning poker to decide how many story points each story was worth depending on the effort needed to reach the acceptance criteria. Anything more than 5 story points was broken down to smaller pieces.

**Creating a Backlog:** The final list of user stories were put in the Backlog. In the Kanban board we made sure that there were no more than 5 user stories at a point in the WIP column.

**Retrospective:** We held regular retrospectives during the workflow along with daily stand ups to understand the roadblocks and problem areas each member in the team was facing and the path to resolve it. We used the smartsheets tool to discuss problems / collaborate and put in the final outcomes of the user stories.

**Agile backlog description:** Using the agile methodology, the team created a backlog with the necessary tasks to complete the analytics project. Since the project was constantly evolving, the initiating and planning tasks were not considered. Additionally, the agile methodology helped reduced the duration of the project by 5 days. The agile backlog included tasks related to “Data gathering and cleaning”, “Data transformation”, “Sentiment analysis”, “Reviews analysis”, “Comparison between the two analysis” and ”Insights generation”. Finally, the agile method has made it possible to divide tasks that were done by several people into single tasks that can now be done by one person. For example: Originally, there was a task to “Compare positive and negative clouds between periods and cities” but now it has been divided into two tasks: The comparison of positive and negative clouds in NYC and London, respectively. (refer Figure 8, 9 and 10)

**Kanban board description:** To build the kanban board, the team used the “Smartsheet platform”. The Kanban board shows the current situation of a project and helps to increase the productivity of the team. In the following screenshot, one can observe the situation of the project in a random moment in time. There is one task that is “At risk”: the “Creation of a PPT/Dashboard to show the insights of the project to our clients”. This has happened because “the comparison of the sentiment analysis with the number of Reviews analysis to understand the "quantity vs quantity" of reviews” and “the creation of a summary to draw conclusions from the analytics” have not been started yet. (See Figure 11)

**Applied Principles (From Manifesto):**

*a) The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.* This has proven to be true during the project. In-person meetings helped the team understand more about the requirements to complete tasks and made it possible to receive feedback from each other. *b) Continuous attention to technical excellence and good design enhances agility.* The team divided the tasks based on the specialties of each member of the team. This increased the productivity of the work done. *c) The best architectures, requirements, and designs emerge from self-organizing teams.* The team members knew their responsibilities and the goals that needed to be accomplished. *d) At regular intervals, the team reflects on how to become more effective then tunes and adjusts its behavior accordingly.* At the end of every meeting, the team members reflected on how to improve the efficiency of resources (time and people) to obtain better results.

# **2. Data Sources Summary:** Our data was compiled from two different sources. The city specific Airbnb data was pulled from Inside Airbnb[[1]](#footnote-2). The decision to choose data from New York and London took much deliberation, but ultimately, we chose those cities based on the availability and quantity of the data. Both New York and London had some of the most plentiful datasets, due to the large number of Airbnb listings in those cities.Our Covid-19 dataset was taken from Our World in Data[[2]](#footnote-3), which is a project performed by the Global Change Data Lab with partnerships from the University of Oxford. OWID has some of the most comprehensive Covid-19 data, including data on cases, deaths, hospitalizations, mortality risk, policy responses, and vaccination rates.

**3. Analysis**

**Cities Considered:** We chose New York and London as our two cities for Analysis since we wanted to gauge how the cultural and geographical differences across these two cities play a factor in determining their respective bookings.

**Analytical Approach:** Our analytical approach involved the following steps. (refer Figure 12)

Doing a sentiment analysis on the reviews for the bookings ii) Analyse the no of reviews to cash some value out of it iii) Compare both to come up with insights that create data-driven insights iv) Running a regression on the factors to gauge the difference in impact of them on bookings across the two different cities that we choose.

**Hypothesis Generation:** To come up with the factors that drive bookings we broke them down into internal factors and external factors (refer Figure 13). The internal factors were i) Features of the property such as Price and Room type ii) Features of the host such as super host, host response rate and host’s profile picture. The external factors were the pandemic and vaccinations drives.

**Covid Effect on Airbnb:** Moving toward the Covid-19 data, we wanted to see if the pandemic had any direct impact on Airbnb’s booking activity. Because Airbnb booking data was not available, we decided to use number of reviews as a proxy variable. Assuming that the number of Airbnb users leaving reviews remains constant over time, this ended up being a decent proxy. To determine Covid-19’s effect on Airbnb, we plotted the number of cases over time side by side with the number of reviews over time. (See Figure 6 and 7). We can see that during the major spikes of Covid-19 cases[[3]](#footnote-4) Airbnb’s activity dropped substantially.

**Sentiment Analysis:** We also wanted to determine if Covid-19 affected the Airbnb’s user experience. This was done by performing a sentiment analysis on reviews from four different time periods: pre-Covid (1/20 – 2/20), Covid (3/20 – 7/20), pre-vaccine distribution (8/20 – 12/20), and post-vaccine distribution (1/21 – present). Splitting reviews by time period and then by positive and negative sentiment, we were looking for trends or differences that would appear across time. There didn’t appear to be any differences between the time periods, but the differences between positive and negative reviews were enlightening. Common keywords used in positive reviews discussed the property and location of the Airbnb listing, while keywords found in negative reviews were often about the host. See Figure 4 and 5**.** This insight led us to the conclusion that users that have a negative experience with Airbnb tend to complain about the host, while those that have a positive experience rarely mention the host.

**City Specific Analysis:** We have analyzed Airbnb activity across different time periods to check for any shifts in the trends for the provided features. Using the number of reviews as a measure of Airbnb activity due to the lack of access to actual booking data, we analyzed these features across two different geographical locations, specifically New York and London. Room Type was the first feature that showed a clear shift in peoples’ choices. Hotel Rooms were the predominant type during the pre-Covid period in New York. However, the activity per listing for hotels decreased as we entered the post-Vaccine period. It is apparent in the graph that shared rooms were becoming less popular after Covid, as people were more interested in private rooms where there was increased activity per listing (Figure 14). Another interesting phenomenon was the price shift after the Covid period, specifically after August of 2021. Listings with prices ranging from 0 to 100 were dominating before covid, while listings priced at 100 to 200 were more popular after that period (Figure 15). Further analysis is still needed to determine the cause behind this phenomenon, since the data does not specify whether it was caused by higher demand raising the price or if people were just willing to spend more during that period. The third feature that affected the activity rate of listings across the four periods is whether hosts had a profile picture or not. There were minimum activity rates for those who had no profile picture during covid and thereafter (Figure 16). Similar trends were found in London, where people were less interested in shared rooms, while entire homes and apartments were becoming more popular post-Covid (Figure 17). However, there was no price shift between the time periods for listings in London (Figure 18). Finally, the third feature showed similar results to New York as less activity was shown for listings where hosts did not have a profile picture (Figure 19).

**Regression:** For analysing how these factors varied for the two cities, we ran a linear regression in python (refer Fig 20) with the factors that were identified in the Step i) and ii) of the analysis. Doing so gave interesting findings which we later incorporated into insights and recommendations:

|  |  |
| --- | --- |
| **New York** | **London** |
| Cleanliness was an important feature for New York | Instant bookableness was an important feature for London |
| Value for money turned out to be a key factor in New York | Accuracy of reviews mattered more for London Airbnbers |
|  | Communication turned out to be a key factor for London |
| Hotel rooms are preferred in New York | Shared rooms are preferred in Londona |

# **4. Proposal for Further Research**

**Recommendations:** From the previous analysis, we have a few recommendations for Airbnb to adapt to best fulfill current customers’ needs and trends. In New York, we would recommend investing in clean, budget-friendly hotel rooms. In London, our recommendation is to invest in instant bookable shared rooms as suggested by the regression model. Communication is key in London and thus it is wise to develop focus in this aspect among hosts. Overall, Airbnb should ensure that hosts have profile pictures as customers are drifting away from listings where host profile pictures are missing. From our sentiment analysis, we saw that location drives positive sentiments while hosts drive negative ones. Investing in host profiles and communication technique would boost activity across Airbnb’s platform.

**Future Steps**:

1. **Urban Vs Rural Activity :** We would like to Analyse how Airbnb locations faired between Urban and Rural areas Pre and Post Covid.
2. **Negative Host Evaluation:** Do a detailed analysis of the Review data to understand what makes a host experience positive or negative and how to counteract a negative review.
3. **Higher Prices:** While we could see a price shift from our analysis, we would like to get more data on the factors that cause it and understand why the demand increases or reduces.
4. **Additional Model Tuning :** Use other models like decision trees and neural networks and understand if they can give us a better response as compared to the regression model we ran.

**Lessons Learned:**

**i) Evolving Requirements:** We followed a flowing Kanban workflow model. Since agile does not depend on a comprehensive set of requirements for the entire project from the beginning and concentrates more on small , complete, independent pieces , it was a lot more accommodating to evolving and changing requirements . The aspect of agile came to our aid as out analysis approach and possible inferences changed as we dived deeper into the data. A waterfall approach would have come to standstill with these changes.

**ii) Resource Allocation**: The Agile approach allows everyone in the team to work on the user stories they want to depending on their bandwidth. Also, the number of user stories in work in progress on a Kanban board is limited to 5. Both of these factors made sure that no one in the team was overworked and negated a need for any kind of resource levelling.

**iii) Cost Optimization and Faster Delivery**: The Agile approach saved us 5 days as compared to the waterfall approach, saving Airbnb considerable amount of money.

**iv) Accountability and Empowerment**: With our approach to Agile , we discussed the acceptance criteria and the story points assigned to each user story via planning poker and team discussions. This approach made sure that each one of us had a say in how defined the scope of our work.This coupled with the fact that we were picking the user stories we wanted to work on on our own and they were not assigned to us, created a system that promoted task accountability and improved the team motivation and commitment to the outcome.There was a personal , invested approach to how we dealt with each user story. This also improved the Quality of work and insights we were able to gather as a apart fo this consulting project.

**References**

<http://insideairbnb.com/get-the-data.html>

2 <https://ourworldindata.org/coronavirus>

3 Major spikes are being defined as the spikes that had the most impact on consumer behavior and public policy (i.e., the first spike, the spike near the end of 2020, the spike of cases caused by the delta variant)

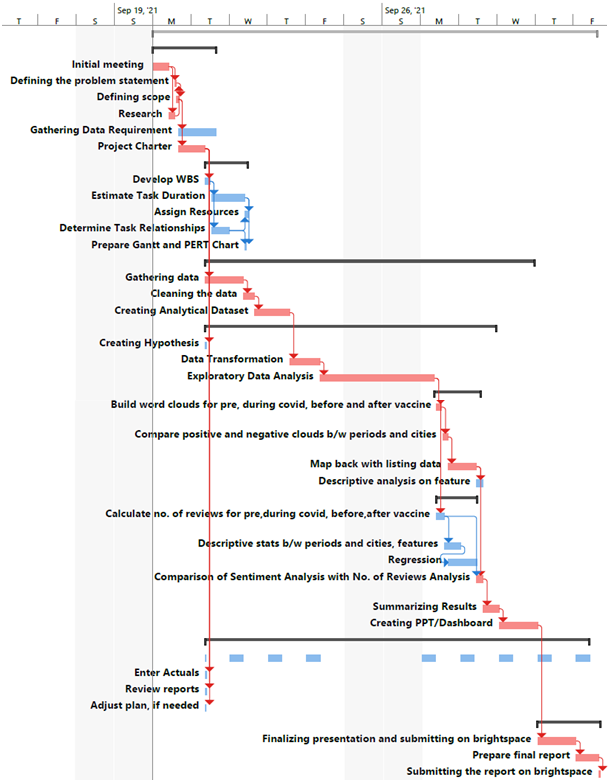
**Appendix**

**Figure 1 Problem Statement**

**Icon, rectangle

Description automatically generated**

**Figure 2**

****

**Figure 3**

Chart, line chart

Description automatically generated

**Figure 4**

Text

Description automatically generated with medium confidence

**Figure 5**

Text

Description automatically generated

**Figure 6**

Chart, line chart

Description automatically generated

**Figure 7**

Chart, line chart

Description automatically generated

**Figure 8: Project Backlog part 1**

Graphical user interface, text

Description automatically generated with medium confidence

**Figure 9: Project backlog part 2**

Graphical user interface, application

Description automatically generated

**Figure 10: User stories (Examples), definition, acceptance criteria and prioritization**

|  |  |  |  |
| --- | --- | --- | --- |
| **User story** | **Definition** | **Acceptance criteria** | **Prioritization** |
| As an analytics consultant, I want to gather Review data to analyze the performance of the hosts and how Covid-19 has changed it. | Review data for different host id’s at a location will be gathered so that it can be used for a sentiment analysis. | Dataset collated for reviews vs different host id’s. | Must have |
| As an analytics consultant, I want to have reliable data related to the hosts and Covid cases so that I can obtain better insights. | Dataset needs to be cleaned. | Dataset cleaned for special characters, categorized and normalized. | Should have |
| As an analytics consultant, I want to do a comparison of the sentiment analysis with the number of Reviews analysis to understand the “quantity vs quantity” of reviews. | Analyze sentiment analysis with reviews analysis to find the features that affect Quality and Quantity of bookings. | Features that affect Quality and Quantity of a booking found | Could have |
| As an analytics consultant, I want to gather data to understand if vaccinated people are more likely to use Airbnb services than those who are not. | Data related to vaccination rates among Airbnb users will be collected to know how it affects the frequency of usage of the platform. | Data about vaccination rates among Airbnb users found. | Won’t have (not this time) |

**Figure 11: Kanban board**

Graphical user interface, application

Description automatically generated

**Figure 12: Analytical Approach**

**A screenshot of a computer screen

Description automatically generated with low confidence**

**Figure 13: Hypothesis**

**A picture containing diagram

Description automatically generated**

**Figure 14: New York Activity Rate vs Room Type**

Graphical user interface

Description automatically generated

**Figure 15: New York Activity Rate vs Price**

Chart, waterfall chart

Description automatically generated

**Figure 16: New York Activity Rate vs Host Profile Picture**

Graphical user interface

Description automatically generated

**Figure 17: London Activity Rate vs Room Type**

Chart, histogram

Description automatically generated

**Figure 18: London Activity Rate vs Price**

A picture containing diagram

Description automatically generated

**Figure 19: London Activity Rate vs Host Profile Picture**

Box and whisker chart

Description automatically generated

**Figure 20: Regression**

***Chart, timeline

Description automatically generated***

1. [↑](#footnote-ref-2)
2. [↑](#footnote-ref-3)
3. [↑](#footnote-ref-4)