IMPACT OF COVID ON AIRBNB

TEAM - 9

ALEXANDER BAKER

AMRIT SINGH

JAIRAJ PARUTHY

SUPRIYA MALLA

VEDANTI GULALKARI

Table of Contents

Acknowledgement

We, the members of Team 9 would like to extend our gratitude towards ***Professor Hossein Ghasemkhani,*** *for teaching us the course "IT Project Management" and giving us the opportunity to work upon this Project. The Project helped us inculcate values of collaboration, teamwork, adherence to tight deadlines and enriched our Project Planning, Data Processing and Data Analysis acumen. The project was also instrumental in increasing team efficiency and taught us practical aspects of planning the scope and executing the tasks efficiently to deliver quality insights in a limited time on a real-time project.*

*We would further like to thank the respected teaching assistant, "****Ilango Guru Muniasamy****" for establishing a robust communication channel and ensuring a smooth course experience.*

We are again grateful for this enriching experience and hope to further augment our Project Management and Analytical skills at Krannert.

Project Management Application

Project management is the supporting pillar of any project on which most of project's success depends. We planned the implementation of our project majorly on two project management tools – MS Project and Atlassian Jira.

Plan Based Approach – Waterfall Model

The project was divided into five summary tasks – initiating tasks, planning tasks, executing tasks, controlling tasks, and closing tasks and three milestones – start project, charter signed and end project. Status task was a recurring task occurring on every Wednesday and Friday. Five resources were assigned to the project, taking the resource levelling into consideration so that no resource has overlapping tasks. The project was hence completed on time within 11.13 days and total cost incurred was $5680.00.

Iterative Approach – Agile Model

To implement agile approach, we leveraged Jira by Atlassian to break our project into 4 sprints and the assigned Epics to the sprints. We divided the epics in the following activities:

- Research data for Airbnb Listings – Sprint 1
- Exploratory Data Analysis on the scraped data – Sprint 2
- Create Analytical processes on explored data – Sprint 3
- Final Report and Presentations - Sprint 4

Post creating the sprints, we broke down our Epics in user stories in the backlog and added the story points corresponding to each task by leveraging the poker system we learnt in IT project Management. We broke down the user stories further down into tasks and held a sprint prioritization meeting to prioritize our activities from backlog into sprint 1 of our JIRA board.

User stories were written in the 'Who/What/Why' format with acceptance criterion added to every user story to help developers visualize the tasks and mark them as completed once the acceptance criterion is met.

Sprint planning meeting were conducted before the start of sprint to help us in deciding the tasks that we wanted to prioritize. Daily standup calls helped us in keeping track of user stories being prioritized by the developers and their status. Sprint review meetings were held towards the end of the sprint where we showcased the completed activities of the sprint. We also

organized Sprint retrospective meetings where we reviewed our sprint to understand ‘What went well’, ‘What did not go well?’ and ‘What we could do better’, we leveraged the sprint retrospective meeting for fine tuning our future sprints and removing any impediments in our project. Dashboard view in Jira helped the team visualize the health of the sprint. Due to the tight timeline, sprint burndown chart helped us provide a clear picture of whether we were on track or were lagging on any activities

We accomplished below project management principles through our analysis project on ‘Effect of COVID-19 on Airbnb:

- **Formal Project Management Structure**

The procedures were planned with flexibility and distributed among all teams’ members according to their schedule. Our project had a project charter signed by all team members, project plan discussed in an initial meeting and designated team member assigned to specific task domain. This helped us successfully prioritize tasks and manage our project.

- **Goals and Outcomes**

The requirements of the project were broken down into precise and measurable outcomes. Each user story was assigned with an acceptance criterion which was tagged to them in Atlassian Jira.

- **Roles and Responsibilities**

The team was assigned roles and responsibilities in terms of planning, executing, testing, controlling and summarization at the initial meeting itself. This helped the team to complete individual tasks and stay on project timeline. The team was also able to contribute to other tasks in case any resource had overlapping tasks. Hence, resource levelling and keeping budget on tasks was possible due to appropriate distribution of roles and responsibilities.

- **Management of Project Changes**

The status meeting was planned as a recurring task. We had retrospective calls planned on every Wednesday and Friday. This helped the team to discuss the tasks that the team did well and the tasks that can be done better. Besides the feedback, the team also discussed the changes that the project might need. In our project, the models and analysis needed careful reconsideration as the data we had at hand was inadequate. We managed to take a business call regarding the same in a status meeting and then responsibilities to be accomplished were redistributed into the respective roles.

- **Risk Management**

The project was monitored to identify, evaluate, and control risks throughout the project timeline. The major risk identified was the constraint of time. We managed the time constraint by following the concept of the iron triangle. We tweaked the scope and cost of the project to balance the time and maintain the quality of the project.

- **Value Delivery Capabilities**

While the planning and designing started as a waterfall approach. As we faced the time constraint, agile approach came in handy and very helpful to collaborate with team members effectively and achieve the project scope on time.

- **Communication**

It is well established that project management is 90% communication. The risks identified and unprecedented changes accounted by the team were effectively taken care of due to astounding team collaboration among, dedication of, and accountability of all team members.

Data Sourcing and Pre-processing

The data was procured from three major data sources – official government website, Inside Airbnb.com and Kaggle.com. The cities under consideration were New York and San Francisco, chosen due to availability of relevant data for years 2019, 2020 and 2021. Major parameter that were exploited for data analysis were listings id, name of the listing, neighbourhood, room type, price, reviews, date, and number of reviews. Data from different sources was collated, mapped and appropriate business rules were applied like filtering null values and categorizing based on listing ids.

Data Analysis and Findings

Descriptive Analysis:

To analysing the data in depth, we first did Exploratory Data Analysis. We focused our analysis on two cities – New York (NY) and San Francisco (SF).

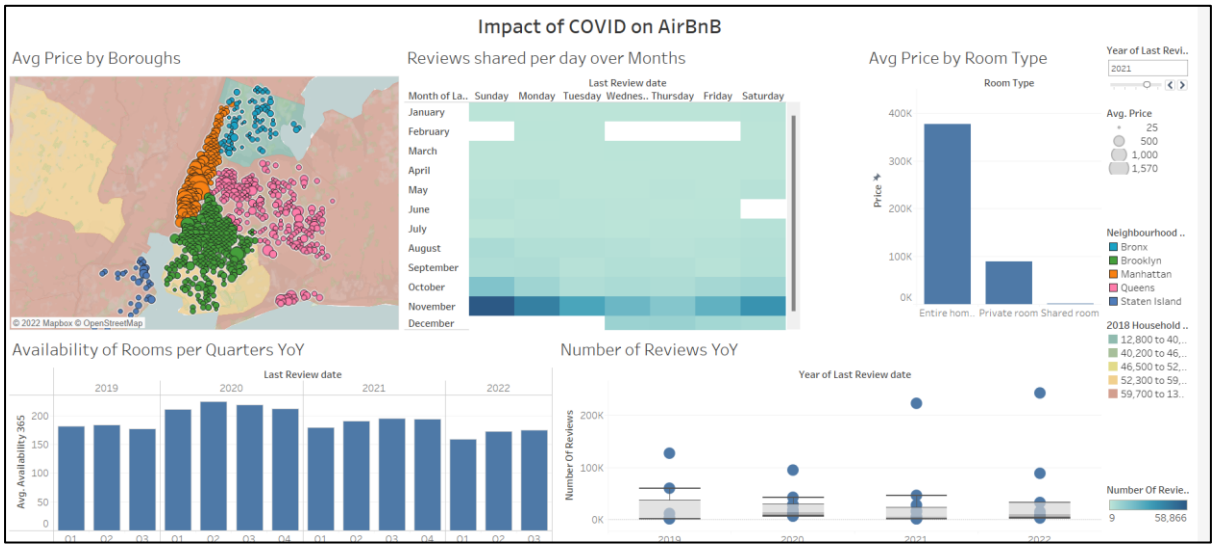
We extracted the COVID cases from early 2020 and overlayed these statistics on the number of reviews Airbnb in NY and SF. We could find a strong correlation as the review volume hit

an all-time low when there was a peak in COVID cases in both these cities. However, we still can't draw a conclusion that number of bookings were also low as we didn't have booking data to corroborate this hypothesis.

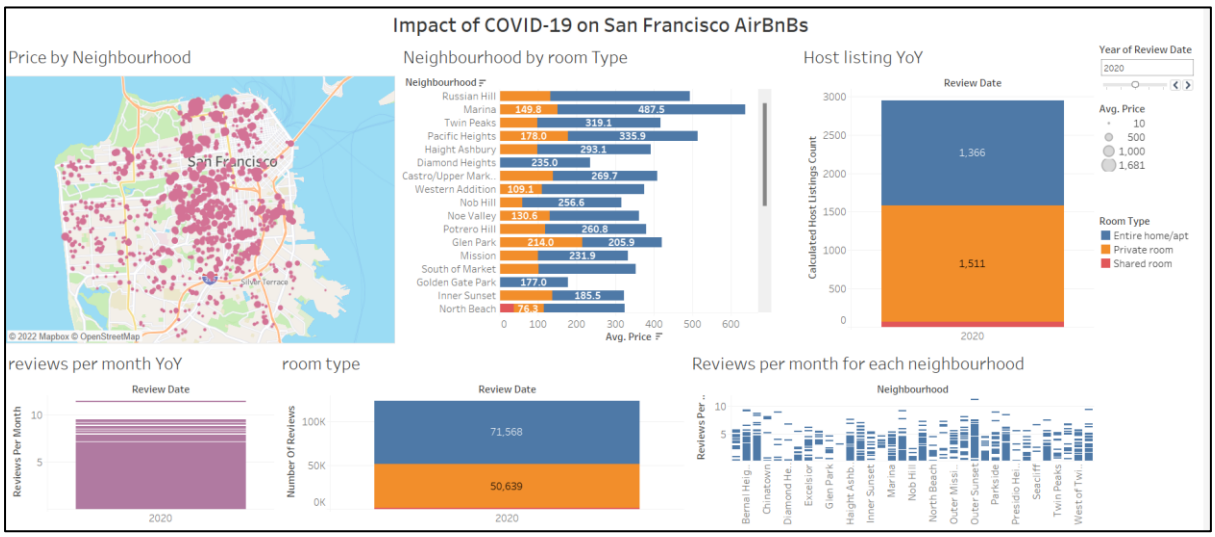
Another interesting aspect, we observed was the change in prices pre covid, during and post covid. And the results were very evident. In SF, we saw the prices drop significantly in 2021 which indicates the aftermath of COVID on the hospitality industry in general.

We created two separate dashboards on Tableau as well to showcase the impact of COVID on different metrics like price, availability of rooms, number of reviews, room types affected the most, host listings, etc. Here are the screenshots of the Tableau dashboards:

New York:



San *Francisco:*



Data Analysis using R:

Using the programming language R, we were able to perform a regression analysis on our data for New York City, and San Francisco. We used the R package “readxl” to input our .xlsx data frames to our environment. Next, using the package “ISLR2” we can fit single-variable generalized linear models to NY and SF.

These two GLMs gives us a preliminary regression equation relating price as a function of the number of reviews. For NY the GLM regression equation was: price=147.9-0.09943(reviews), and for SF the equation was: price=232.6-0.3119(reviews).

But in ISLR2, we can account for more than just a single independent variable, we can model multiple regressions quite easily. Doing this for NY with three independent variables, and SF with two independent variables, we get these final multiple regressions from our GLMs.

GLM multiple linear regression for NY: Price = 147.46322 - 0.09009(number_of_reviews) - 3.46055(reviews_per_month) + 0.06215(availability_365)

GLM multiple linear regression for SF: price = 279.6542 - 0.4322(number_of_reviews) - 3.7953(Minimum_Number_Of_Nights)

These models allow us to predict the price of a unit with statistical confidence. We can rely on these models, because their leverage analysis presents no influential points, and the residuals analysis shows limited heteroskedasticity.

We can utilize this data analysis to inquire about specific units, such for one in NY where we know the three independent variables. If we know the number of reviews, how many per month, and what the yearly availability is, we can use those three attributes to make a prediction about the expected price of the unit.

This is an advantage for any real estate company, allowing them to model how they expect the market to behave with regards to a given unit. Also allowing as a failsafe, to check if the unit in question is overpriced according to the regression model. If the model predicts a significantly lower price, an inquiry can be made into the property to determine if the unit is competitively priced.

Proposal for further research

Given the limited time, and availability of resources, we had to limit our scope in order to ensure the quality of deliverables which concurs with the principles of Iron Triangle. The scope got limited owing to one of the major challenges that we faced during data gathering and pre-processing stage which was 'unavailability or lack of access to customer level data' of AirBnb. Unavailability of the bookings data prevented us from establishing accurate relationship between covid trends, the no of reviews being posted and the pandemic's impact on revenues of AirBnB. Therefore, one the most essential steps as part of further research is to get in touch with official Airbnb official sources and channels and try to procure bookings level data. We could then use that data to match the customer reviews responses with their ticket size to cluster the consumer in different segments like, "Luxury Loving Customers", "Influencers", "Business Travelers", or "Casual Travelers". This segmentation study could have greatly helped AirBnB to enact appropriate Business strategies and marketing tactics to provide an impetus to growth and recovery post Covid-pandemic.

We would further utilize the consumer bookings data to establish correlation with other features like room-type, reviews per room, neighborhood prices etc, to get the pricing trends pre-pandemic, during pandemic and post-pandemic. Furthermore, these features can be incorporated to leverage ensemble of models like Neural Network, Random Forests, XGBoost, etc for the purpose of predictive modelling. Lastly, the previously mentioned models would be trained and tested on different data-sets during the three different periods, Pre-Covid, During-Covid, and Post-Covid so as to predict pricing trends and consumer behavior during similar times. This shall help Airbnb to enact and implement contingency measures as well as diversify their business operations in case a Pandemic like situation arises again.

Appendix:

Exhibit 1: MS Project

MPP File for MS Project Plan: https://purdue0-my.sharepoint.com/personal/vgulalka_purdue_edu/_layouts/15/onedrive.aspx?id=%2Fpersonal%2Fvgulalka%5Fpurdue%5Fedu%2FDocuments%2FMicrosoft%20Teams%20Chat%20Files%2FMGMT585%5FTeam9%5FMS%5FProject%2Empp&parent=%2Fpersonal%2Fvgulalka%5Fpurdue%5Fedu%2FDocuments%2FMicrosoft%20Teams%20Chat%20Files&ga=1

MGMT585_Team9_MPP

	Task Name	Duration	Start	Finish	Predecessors	Resource Names
0	**MGMT585_Team9_MS_Project**	**11.13 days**	**Mon 11/21/2**	**Tue 12/6/22**		
1	1 Start Project	0 days	Mon 11/21/22	Mon 11/21/22		
2	**2 Initiating Task**	**3 days**	**Mon 11/21/22**	**Wed 11/23/22**	**1**	
3	2.1 Initial Meeting to discuss project scope	0.5 days	Mon 11/21/22	Mon 11/21/22	1	Alexander Baker,Amrit Singh,Jairaj Paruthy,Supriya Malla,Vedanti Gulalkari
4	2.2 Research on AirBnB listings	0.5 days	Mon 11/21/22	Mon 11/21/22	3	Amrit Singh[50%],Jairaj Paruthy[50%]
5	2.3 Perform Data Cleaning and Preprocesing	1 day	Tue 11/22/22	Tue 11/22/22	4	Alexander Baker
6	2.4 Draft User Stories	1 day	Tue 11/22/22	Tue 11/22/22	4	Amrit Singh
7	2.5 Develop Project Charter	1 day	Wed 11/23/22	Wed 11/23/22	6	Supriya Malla
8	2.6 Charter Signed	0 days	Wed 11/23/22	Wed 11/23/22	7	Jairaj Paruthy[50%]
9	**3 Planning Task**	**1 day**	**Thu 11/24/22**	**Thu 11/24/22**	**8**	
10	3.1 Develop WBS	0.5 days	Thu 11/24/22	Thu 11/24/22	8	Vedanti Gulalkari
11	3.2 Estimate Task Duration	0.5 days	Thu 11/24/22	Thu 11/24/22	10	Supriya Malla[50%]
12	**4 Executing Tasks**	**3.5 days**	**Fri 11/25/22**	**Wed 11/30/22**	**11**	
13	4.1 Exploratory Data Analysis	1 day	Fri 11/25/22	Fri 11/25/22	11	Jairaj Paruthy[50%],Amrit Singh[50%]
14	4.2 Design Hypothesis	0.5 days	Mon 11/28/22	Mon 11/28/22	13	Vedanti Gulalkari[50%]
15	4.3 Model Building	1 day	Mon 11/28/22	Tue 11/29/22	14	Alexander Baker
16	4.4 Model Training and Validation	1 day	Tue 11/29/22	Wed 11/30/22	15	Alexander Baker[50%]
17	**5 Controlling Tasks**	**6.13 days**	**Thu 11/24/22**	**Fri 12/2/22**	**8**	
18	**5.1 Status Report**	**5.13 days**	**Fri 11/25/22**	**Fri 12/2/22**	**8**	
19	5.1.1 Status Report 1	1 hr	Fri 11/25/22	Fri 11/25/22		Amrit Singh[20%]
20	5.1.2 Status Report 2	1 hr	Wed 11/30/22	Wed 11/30/22		Supriya Malla[20%]
21	5.1.3 Status Report 3	1 hr	Fri 12/2/22	Fri 12/2/22		Vedanti Gulalkari[20%]
22	5.2 Enter Actuals	0.5 days	Thu 11/24/22	Thu 11/24/22	8	Jairaj Paruthy[50%]
23	5.3 Review Report	0.5 days	Thu 11/24/22	Thu 11/24/22	8	Amrit Singh[50%]
24	5.4 Adjust Plan, if needed	0.5 days	Thu 11/24/22	Thu 11/24/22	8	Vedanti Gulalkari
25	**6 Closing Tasks**	**3.63 days**	**Wed 11/30/22**	**Tue 12/6/22**	**16**	
26	6.1 Prepare Inferences	1 day	Wed 11/30/22	Thu 12/1/22	16	Amrit Singh
27	6.2 Prepare Project Final Report	0.5 days	Thu 12/1/22	Thu 12/1/22	26	Supriya Malla
28	6.3 Prepare Project Presentation	1 day	Fri 12/2/22	Fri 12/2/22	27	Vedanti Gulalkari
29	6.4 Prepare Lessons Learned	1 day	Mon 12/5/22	Mon 12/5/22	28	Alexander Baker
30	6.5 Deliver Final Project	1 hr	Tue 12/6/22	Tue 12/6/22	29	Jairaj Paruthy
31	6.6 End Project	0 days	Tue 12/6/22	Tue 12/6/22	30	

Figure 1. MS Project Plan and Resource Allocation

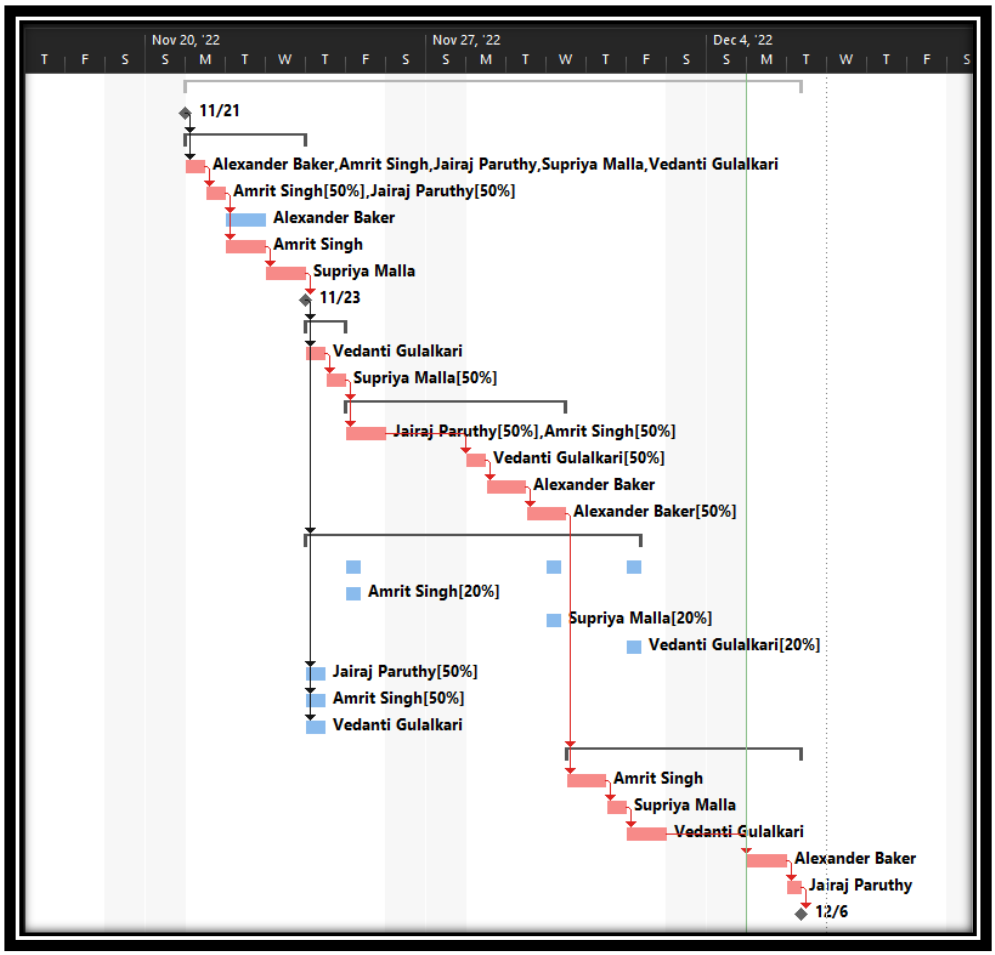
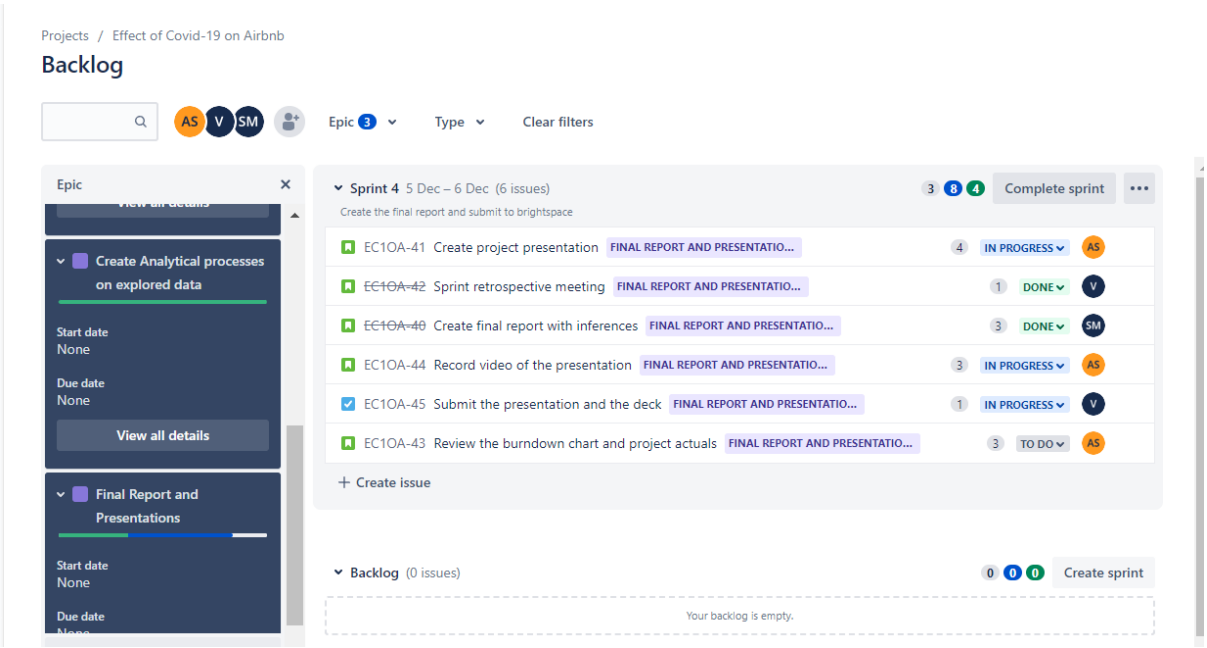


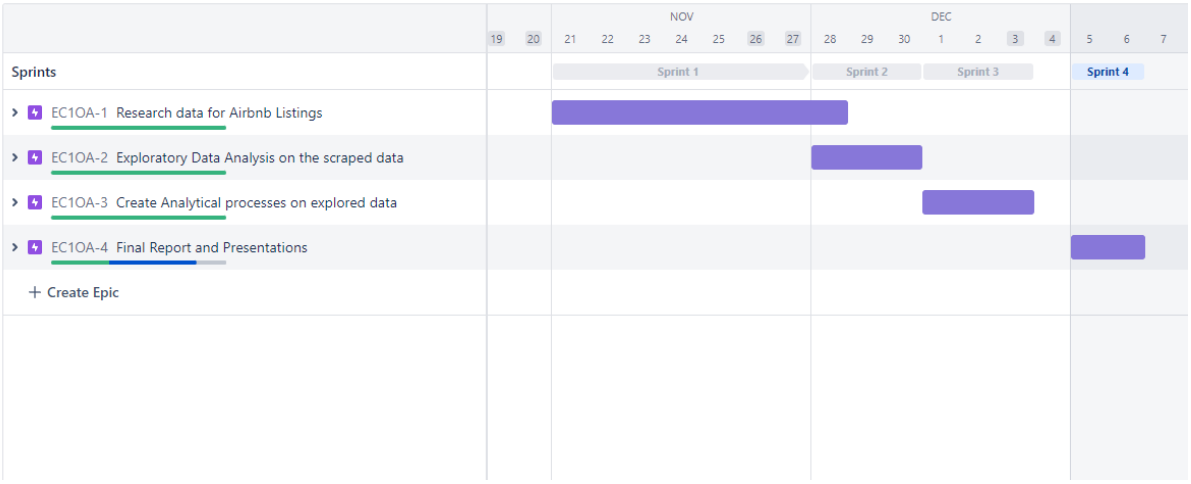
Figure 2. Gantt Chart for the proposed MS Project Plan

Exhibit 2: JIRA Board

Backlog view of JIRA:



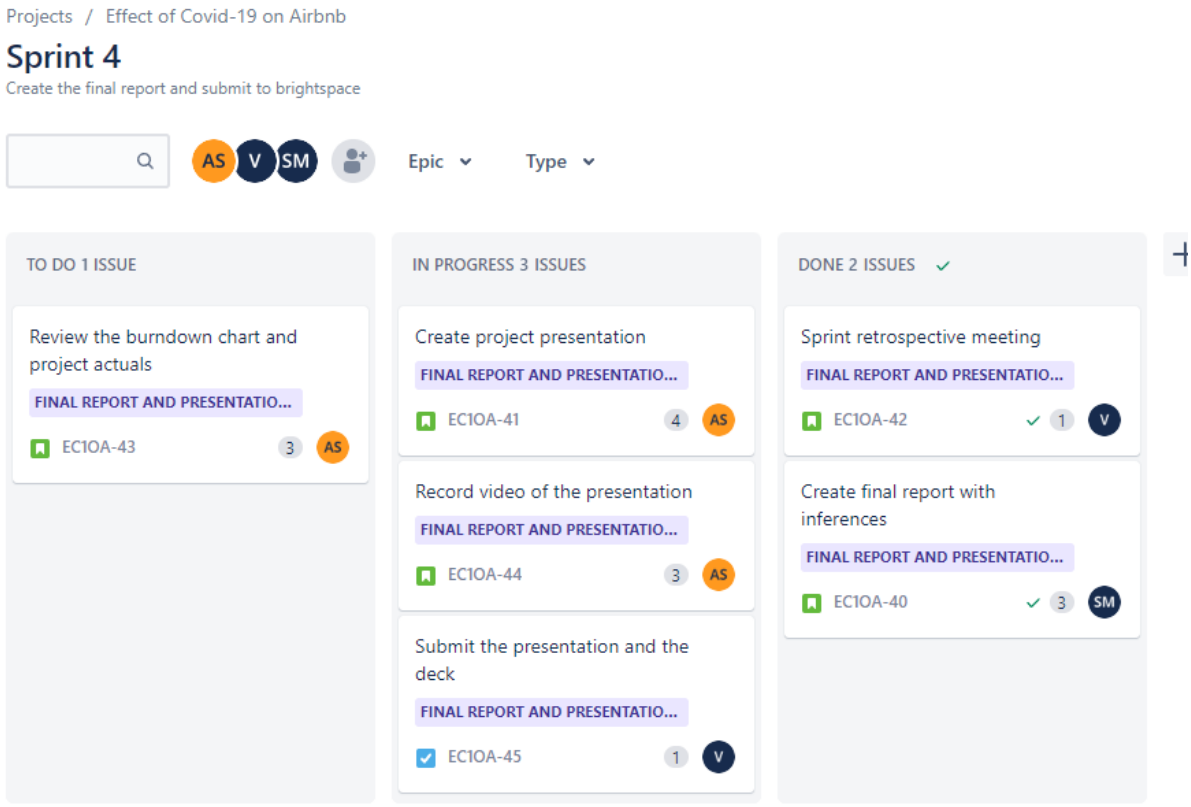
Road map view of JIRA board:



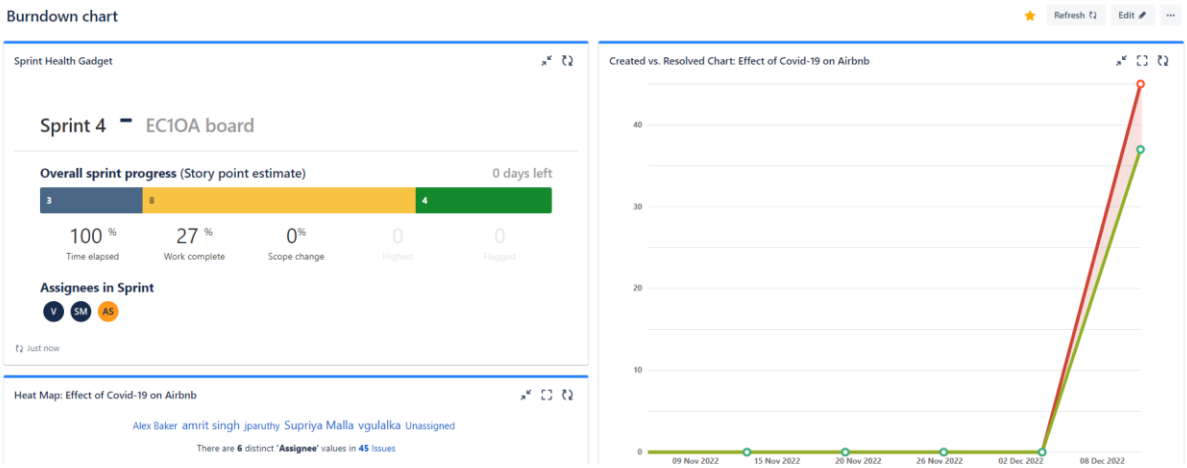
Kanban board view of JIRA:

Link to Kanban board:

https://amritsingh.atlassian.net/jira/software/projects/EC1OA/boards/1/backlog?issueParent=10001%2C10002%2C10003&selectedIssue=EC1OA-4



Dashboard view for tracking sprint health:



User story and Acceptance criterion:

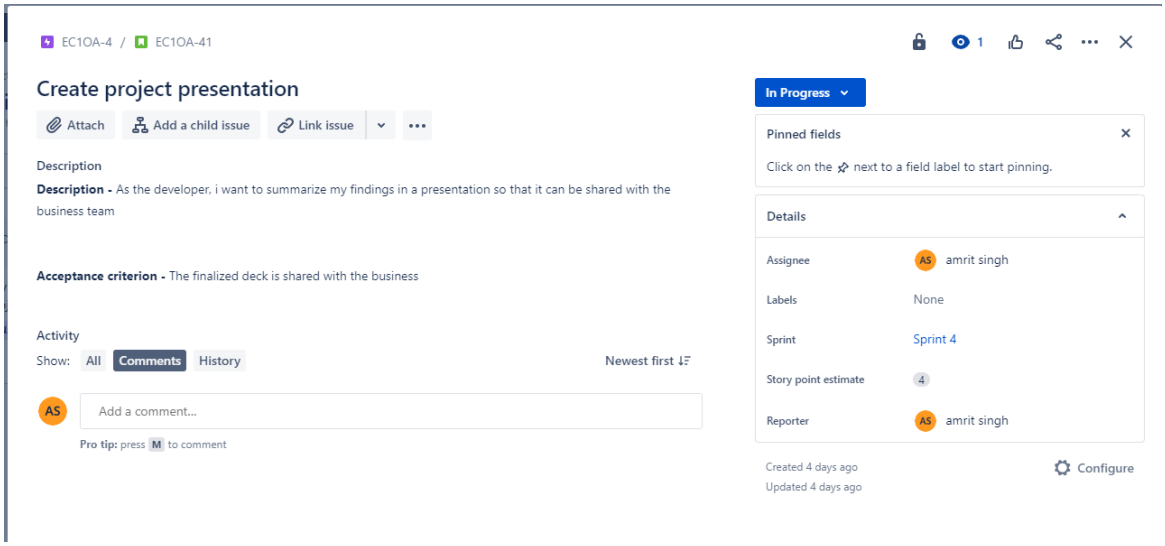


Exhibit 3: Tableau Dashboards

AirBnBs in San Francisco COVID Impact:

https://public.tableau.com/views/ImpactofCOVID-19onSFAirBnB/Dashboard1?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

AirBnBs in New York COVID Impact:

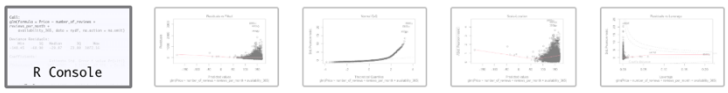
https://public.tableau.com/views/ImpactofCOVIDonNYAirBnB/ImpactofCOVIDonAirBnB?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

Exhibit 4: Data Analysis in R

NY multiple regression code

````
Fitting a Generalized linear model for New York:
```{r}
fit.glm <- glm(Price ~ number_of_reviews, data=nydf, na.action = na.omit)
summary(fit.glm)
plot(fit.glm)
?glm
```
GLM regression equation: price = 147.9 - 0.09943(reviews)


```{r}
fit.glm <- glm(Price ~ number_of_reviews + reviews_per_month + availability_365, data=nydf, na.action = na.omit)
summary(fit.glm)
plot(fit.glm)
```
````




```
Call:
glm(formula = Price ~ number_of_reviews + reviews_per_month +
    availability_365, data = nydf, na.action = na.omit)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-148.45   -68.90   -29.87    23.80  3072.16

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       147.46322    3.11596  47.325  < 2e-16 ***
number_of_reviews  -0.09009    0.01911  -4.713 2.47e-06 ***
reviews_per_month  -3.46055    0.65619  -5.274 1.37e-07 ***
availability_365    0.06215    0.01144   5.432 5.70e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

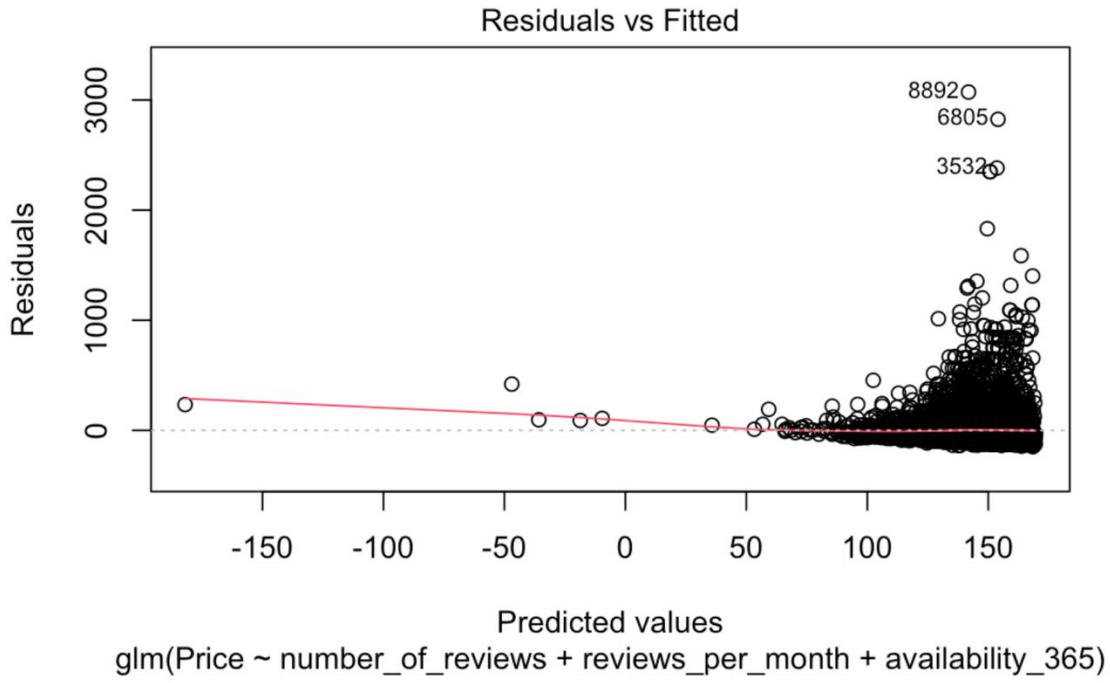
(Dispersion parameter for gaussian family taken to be 19160.34)

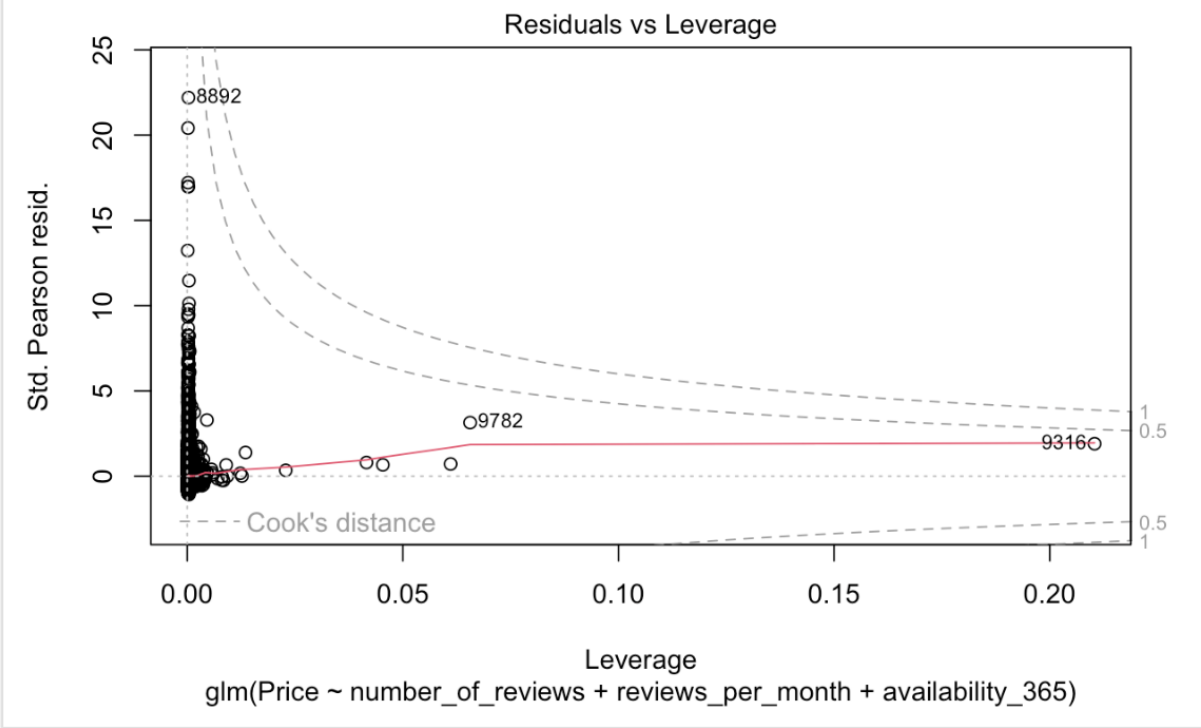
    Null deviance: 190214330  on 9818  degrees of freedom
Residual deviance: 188058765  on 9815  degrees of freedom
AIC: 124692

Number of Fisher Scoring iterations: 2
```

````
multiple linear regression for NY: Price = 147.46322 - 0.09009(number_of_reviews) - 3.46055(reviews_per_month) + 0.06215(availability_365)
````

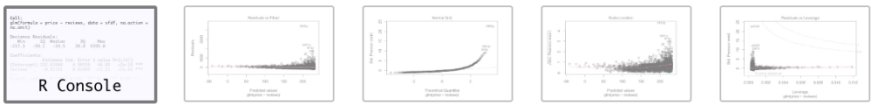
NY leverage and residuals plots





SF multiple regression code

```
Fitting a Generalized linear model for San Francisco:
```{r}
fit.glm <- glm(price ~ reviews, data=sfdf, na.action = na.omit)
summary(fit.glm)
plot(fit.glm)
```
```




```
Call:
glm(formula = price ~ reviews, data = sfdf, na.action = na.omit)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-217.3   -99.1   -39.5    30.8  4395.0

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 232.62488    4.96556   46.85   <2e-16 ***
reviews      -0.31191    0.02809  -11.11   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

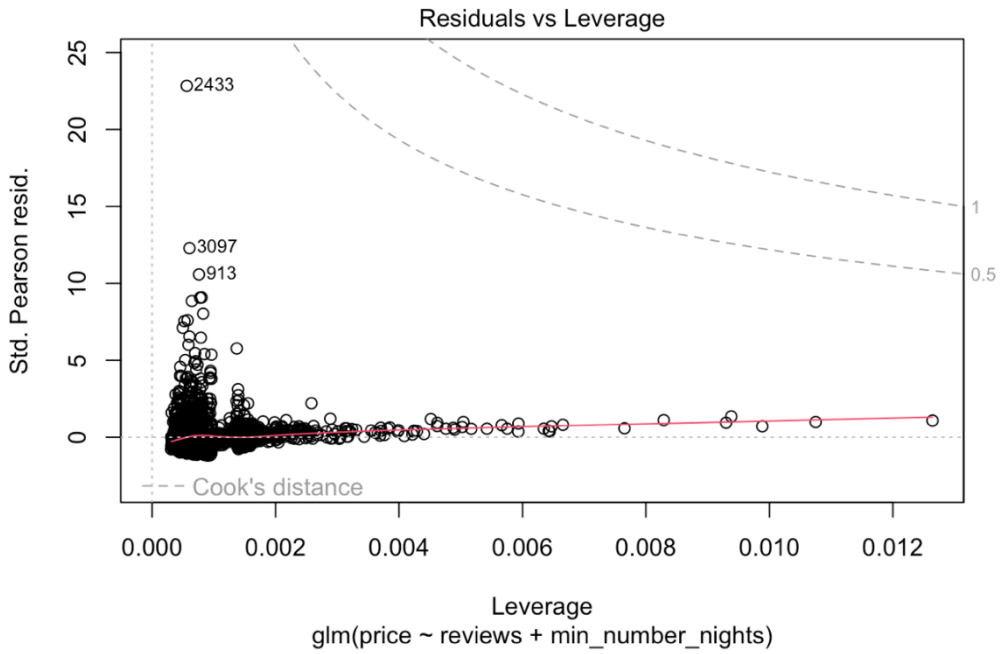
(Dispersion parameter for gaussian family taken to be 38241.05)

    Null deviance: 129916782  on 3275  degrees of freedom
Residual deviance: 125201196  on 3274  degrees of freedom
AIC: 43868

Number of Fisher Scoring iterations: 2
```

```
multiple linear regression for SF: price = 279.6542 - 0.4322(number_of_reviews) - 3.7953(Minimum_Number_Of_Nights)
```

SF leverage and residuals plots



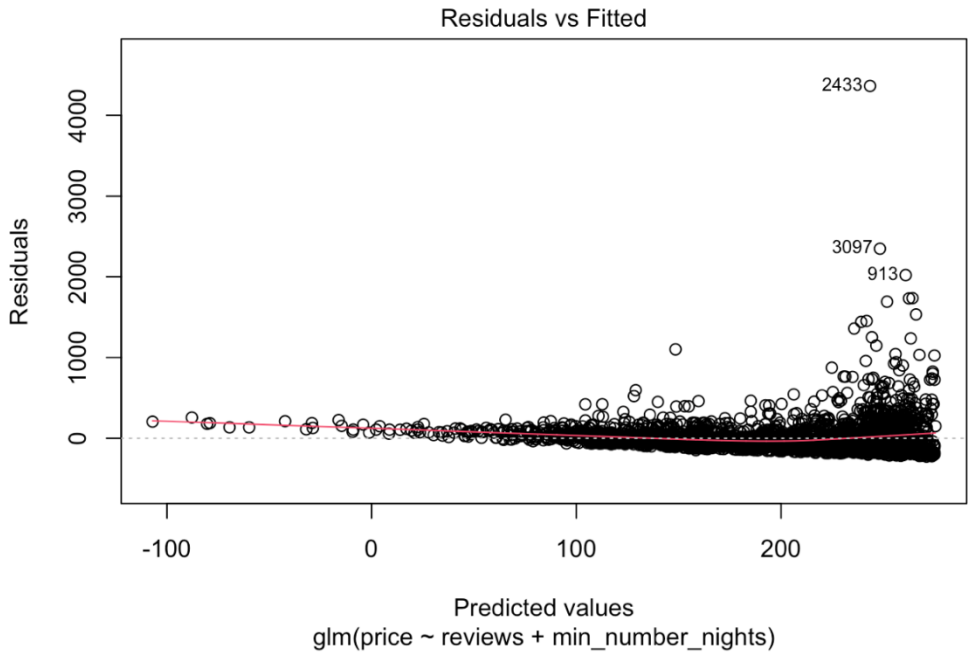


Exhibit 5: Presentation Link

PPT link to the Project Presentation:

https://purdue0-my.sharepoint.com/:p:/g/personal/malla2_purdue_edu/EeKm64iFQ4hEoncN8Ko7d7MBgOLqmR_fPCenfH9WPwGbIQ

ITPM Final Project Presentation.pptx