AIRBNB RATING SYSTEM

A Hypothesis Testing Exercise on What Driving Airbnb High Ratings

Andrew Wong

ABOUT ME

Based in Melbourne, Australia. I have worked / lived in 14 countries in Europe and Asia.

A high-performing commercial technologist and solution delivery leader with a track record in working with data science, mobile, digital and software engineering program budget up to AUD$50 million.

I have specialist data science domain in the followings

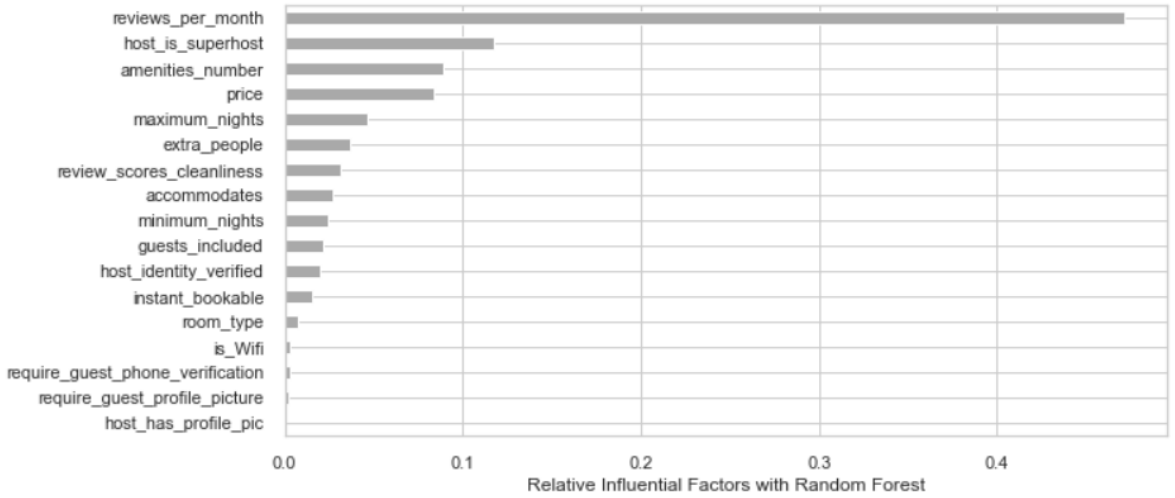
- ***Agile Development:*** *Rapid iteration of data science projects through small, incremental value delivery.*
- ***Product Development:*** *Developing minimum viable data science / machine learning product (MVP) through Design Thinking and Design Sprinting.*
- ***Industry Verticals:*** *Start-ups, MedTech, EdTech, FinTech*

BUSINESS PROBLEM

A Further Examination of Airbnb Rating Systems - Influential Factors

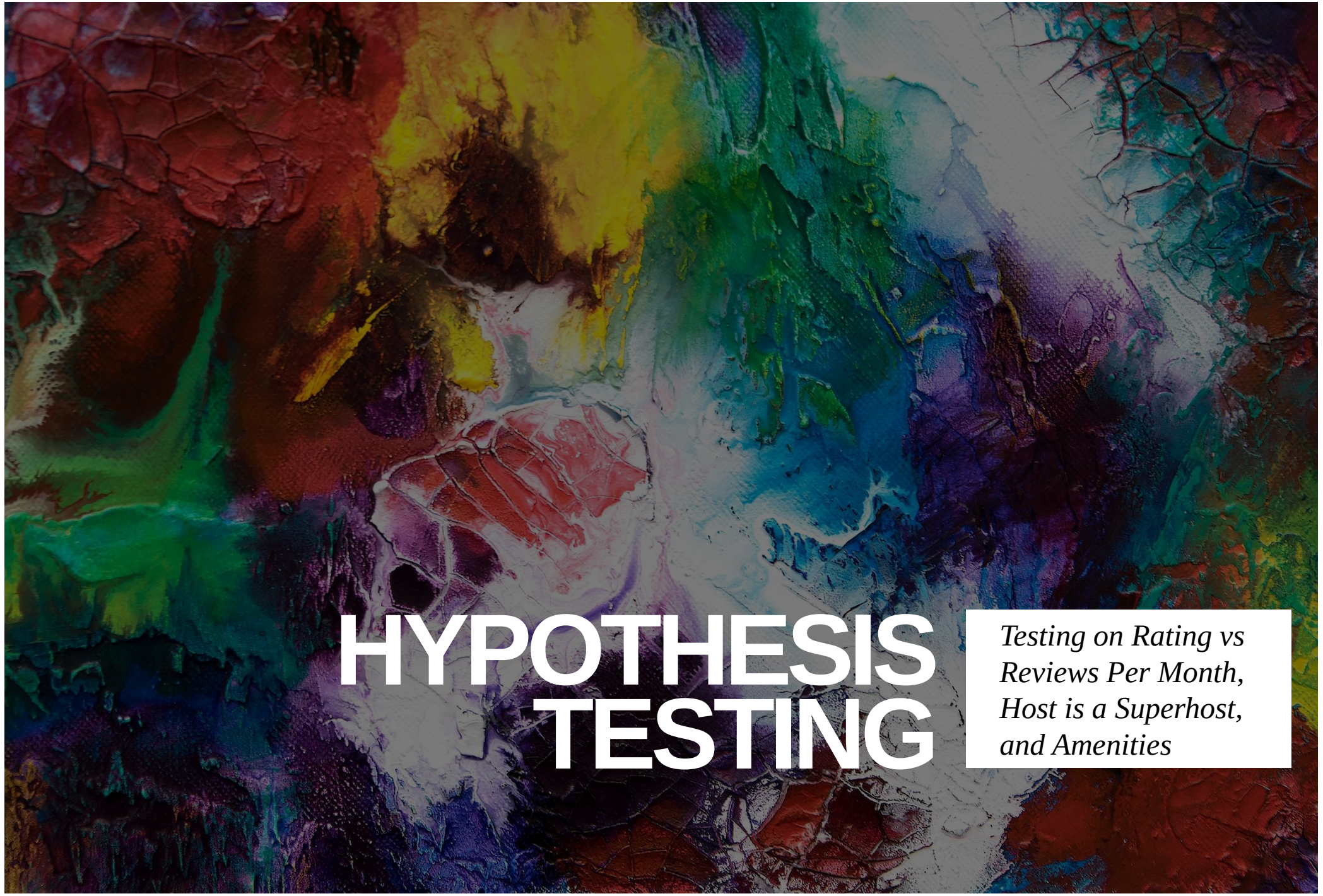
RELATIVE INFLUENTIAL FACTORS OF AIRBNB RATING

Top three influential factors of high host rating: reviews per month, host is a superhost, and amenities



Business Problem:

The current rating system are influenced by more than 20 plus factors (features). I have derived the top few factors through machine learning modelling (Random Forrest). A more nuance analysis of these top influential factors will help to inform business strategic decision on where to invest next.


HYPOTHESIS
TESTING
Testing on Rating vs
Reviews Per Month,
Host is a Superhost,
and Amenities

HYPOTHESIS TESTING RESULT - 1

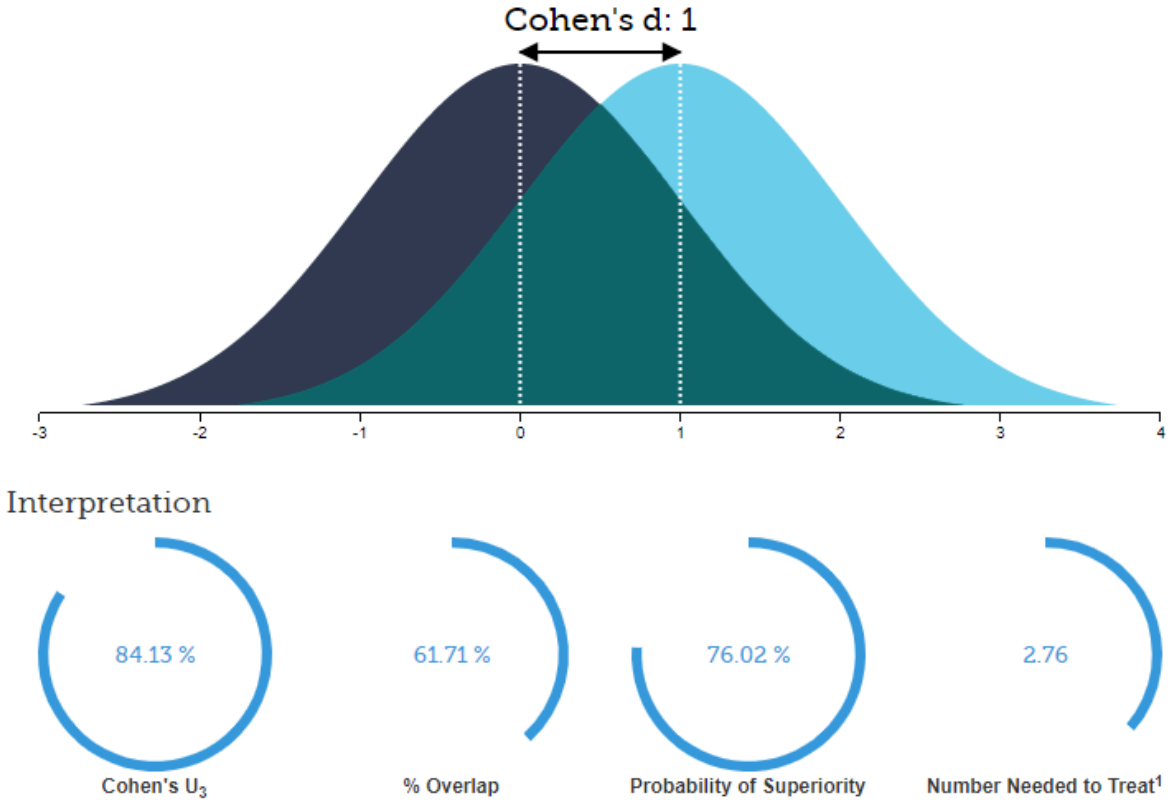
Rating vs Reviews per Month

Outcome Hypothesis Test 1 Average Reviews Per Month and High Rating Given to Airbnb Host

> **T-Test:** Statistically significant with a p < .001. Therefore, we reject the null hypothesis, and accept the alternative hypothesis.

> **Welch's T-Test:** Statistically significant with a p < .001. Therefore, we reject the null hypothesis, and accept the alternative hypothesis.

> **Cohen Effect Size:** High effect size with a 1.03.



Interactive Cohen's D credit to Kristoffer Magnusson (https://rpsychologist.com/d3/cohend/)

HYPOTHESIS TESTING RESULT - 2

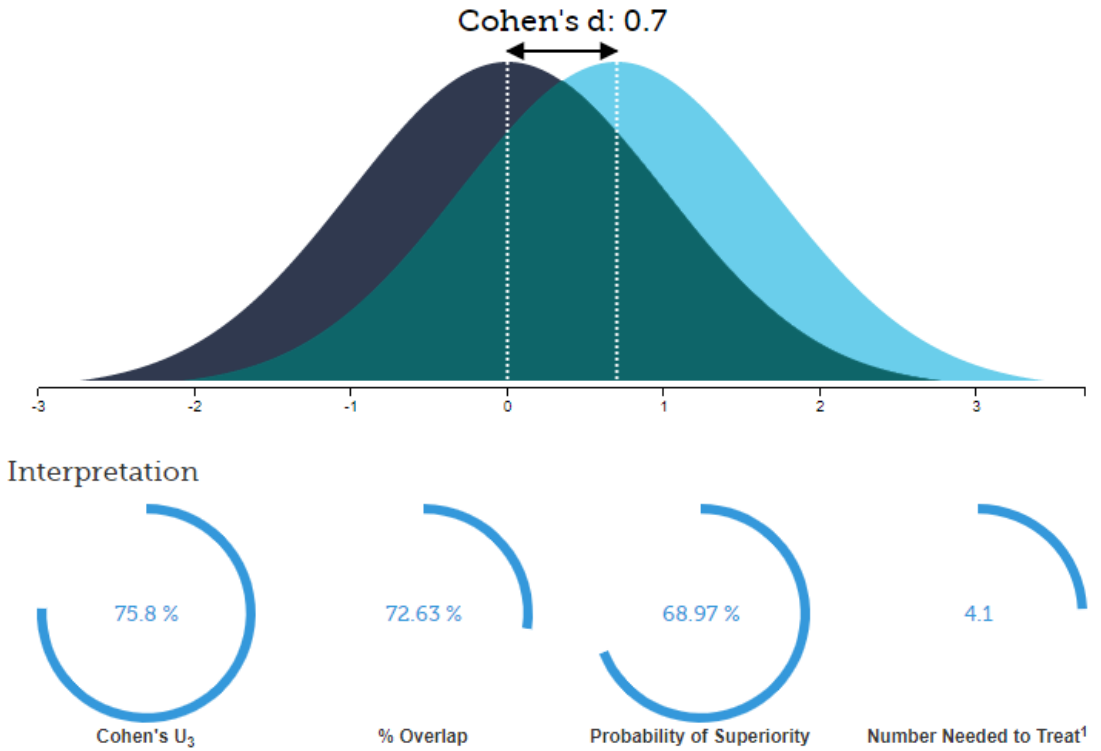
Rating vs Number of Amenities

Outcome Hypothesis Test 2 Average Number of Amenities and High Rating Given to Airbnb Host

> **T-Test:** Statistically significant with a $p < .001$. Therefore, we reject the null hypothesis, and accept the alternative hypothesis.

> **Welch's T-Test:** Statistically significant with a $p < .001$. Therefore, we reject the null hypothesis, and accept the alternative hypothesis.

> **Cohen Effect Size:** Medium effect size with a 0.7.



Interactive Cohen's D credit to Kristoffer Magnusson (https://rpsychologist.com/d3/cohend/)

HYPOTHESIS TESTING RESULT - 3

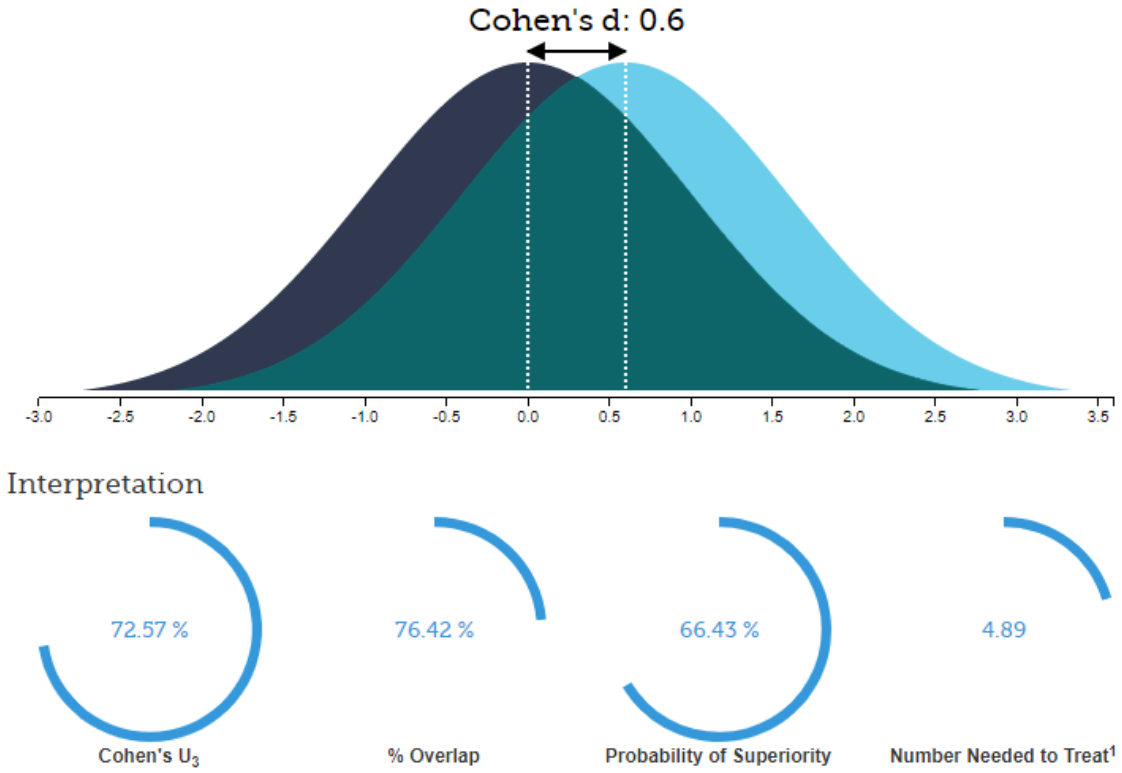
Rating vs Host is a Superhost

Outcome Hypothesis Test 3 Host is a Superhost and High Rating Given to Airbnb Host

> **T-Test:** Statistically significant with a p < .001. Therefore, we reject the null hypothesis, and accept the alternative hypothesis.

> **Welch's T-Test:** Statistically significant with a p < .001. Therefore, we reject the null hypothesis, and accept the alternative hypothesis.

> **Cohen Effect Size:** Medium effect size with a 0.6.



Interactive Cohen's D credit to Kristoffer Magnusson (https://rpsychologist.com/d3/cohend/)

RECOMMENDATIONS

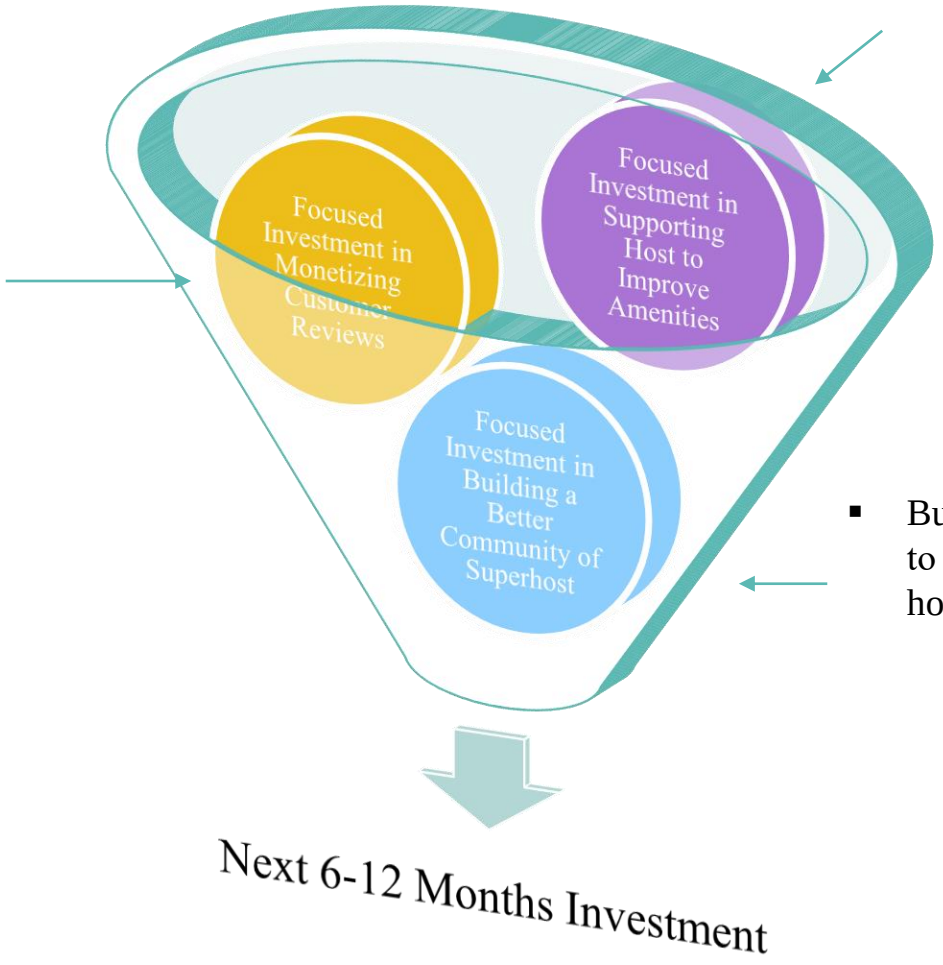
What Can We Do With These Hypothesis Testing Results - For Data Scientists and Business Decision-Makers

RECOMMENDATIONS FOR BUSINESS DECISION-MAKERS (AIRBNB EXECUTIVE)

80 / 20 Percent Rules of Investing – Start Investing Wisely

- Support host to focus on enhancing high-value amenities

- Building better sentiment analysis engine
- Building a more interactive customer reviews systems



- Build a Superhost support group to capitalise on Superhosts' know-how

RECOMMENDATIONS FOR DATA SCIENTISTS

Advancing the discussion between Effect Size and P=Value

- *Discuss:* The p-value position is under threat
- *Discuss the difference:* What is the relationship between 'effect size' and 'significance'?
- *Build case studies:* Statistical significance does not tell you the most important thing, unlike the size of the effect

LIMITATIONS AND FUTURE SCOPE

Next Chapter

LIMITATIONS

A short hypothesis testing exercise

Business recommendations require further customer validation, not just rely on hypothesis testing and machine learning modeling

FUTURE SCOPE

Run other specific customer reviews rating score

Run a longitude study with data from 2-3 years

THANK YOU

Andrew Wong

+61 416527928

andrewwongls@outlook.com

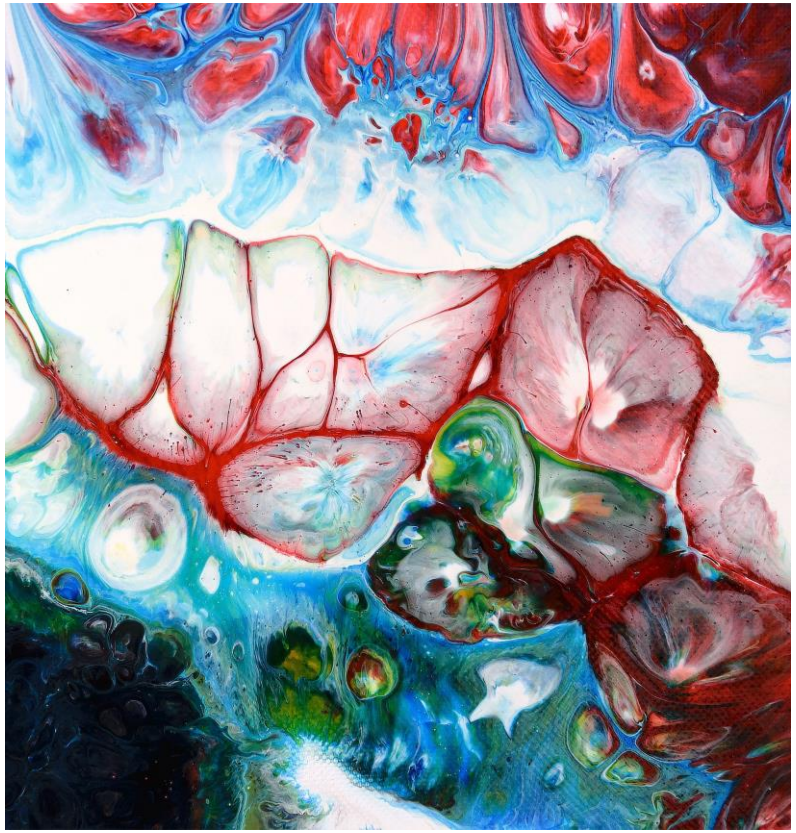
LinkedIn: linkedin.com/in/andrewongls

GitHub: github.com/andrewongls

Blog: medium.com/human-science-ai

APPENDIX

References, additional analysis and figures



REFERENCES

Hypothesis Testing, T-Test, Welch's, and Cohen's D

Hypothesis testing is used to infer the result of a hypothesis performed on sample data from a larger population. The test tells the analyst whether or not his primary hypothesis is true.

A **t-test** is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related in certain features. **t-value** is a value which is used in t-test. The t-test helps to find out the difference in the average of one or more than one (two) population distributions. With every t-value, there is p-value that helps you to make the process of finding the difference easier. A t-value is the relative error difference in contrast to the null hypothesis.

Welch's t-test, or unequal variances t-test, is a two-sample location test which is used to test the hypothesis that two populations have equal means. Welch's t-test is more robust than Student's t-test and maintains type I error rates close to nominal for unequal variances and for unequal sample sizes under normality.

Furthermore, the power of Welch's t-test comes close to that of Student's t-test, even when the population variances are equal and sample sizes are balanced.

Cohen's D is one of the most common ways to measure effect size. An effect size is how large an effect of something is. For example, medication A has a better effect than medication B.

REFERENCES

P-value, and Cohen's D

P-value	Interpretation
Over 0.1	No evidence that the null hypothesis does not hold
Between 0.05 and 0.1	Very weak evidence that the null hypothesis does not hold
Between 0.01 and 0.05	Moderately strong evidence that the null hypothesis does not hold
Under 0.01	Strong evidence that the null hypothesis does not hold
How to read the result: A p-value, is the statistical significance of a measurement in how correct a statistical evidence part, is P-values are most likely to be near 0 if the alternative hypothesis holds If a data set gives rise to a p-value of say 0.0023, we can state that the probability of getting a data set with such a low p-value is only 0.0023 if H0 is true. Since such a low p-value is so unlikely, the data give strong evidence that H0 does not hold. Of course, we may be wrong. A p-value of 0.0023 could arise when either H0 or HA holds. However it is unlikely when H0 is true and more likely when HA is true.	

Effect Size (Cohen's D)	Interpretation
0.2	Considered a 'small' effect
0.5	Considered a 'medium' effect
0.8	Considered a 'large' effect
How to read the result: This means that if two groups' means don't differ by 0.2 standard deviations or more, the difference is trivial, even if it is statistically significant.	