01

Priya Seejoor
Capstone | 2020

Ad Fraud

70 %

Active Mobile Devices Nationwide

3 Billion

Clicks Per Day

90 %

Potentially Fraudelant Data

Data Set

FROM KAGGLE:

TalkingData AdTracking Fraud Detection
November 6, 2017 4pm - November 9, 2017 3pm

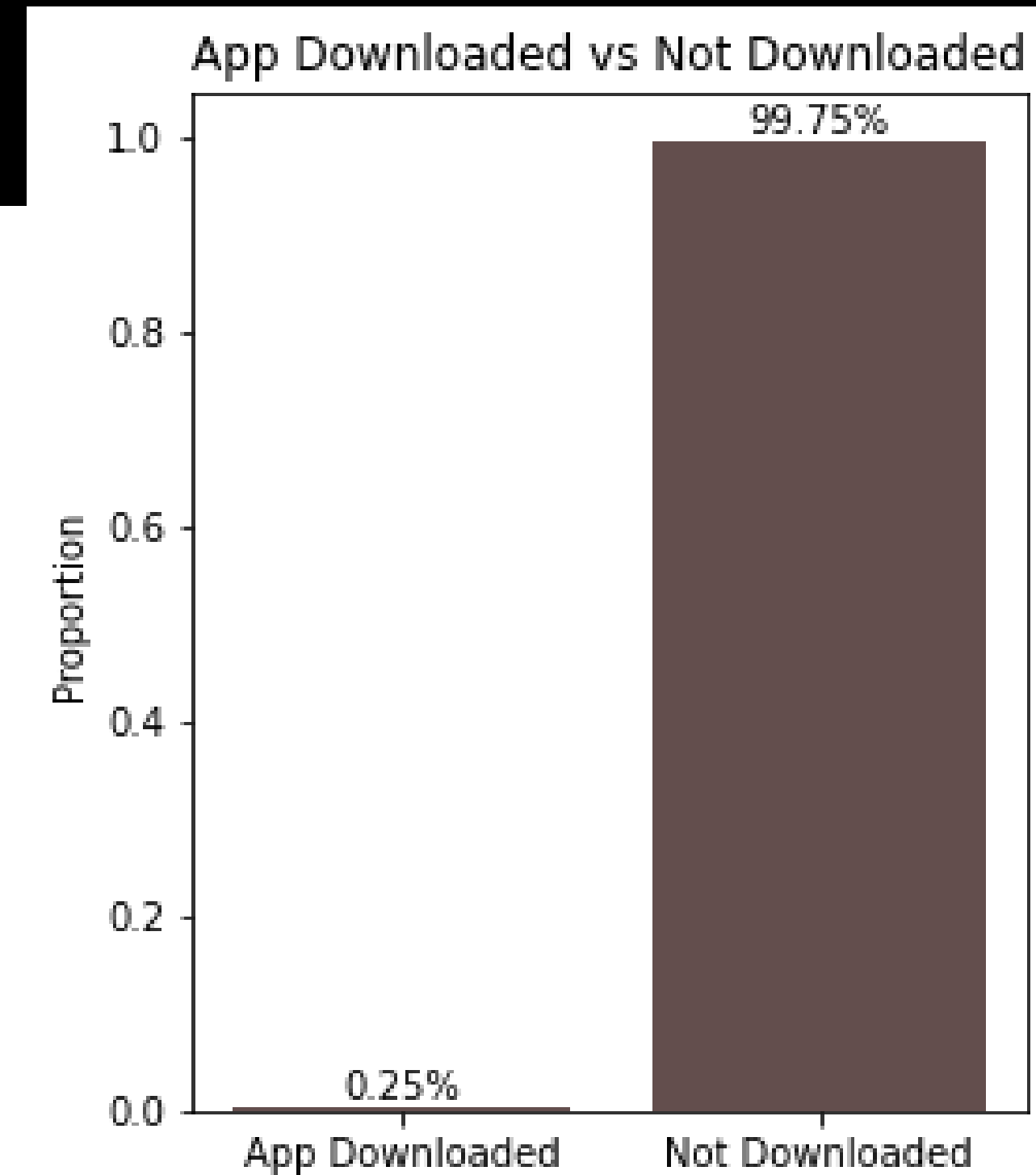## H_o

Rate of fraud at night = rate of fraud during the day

H_a

Rate of fraud at night ≠ rate of fraud during the day

04

WHAT CAN WE INFER?

First Glance at the dataset
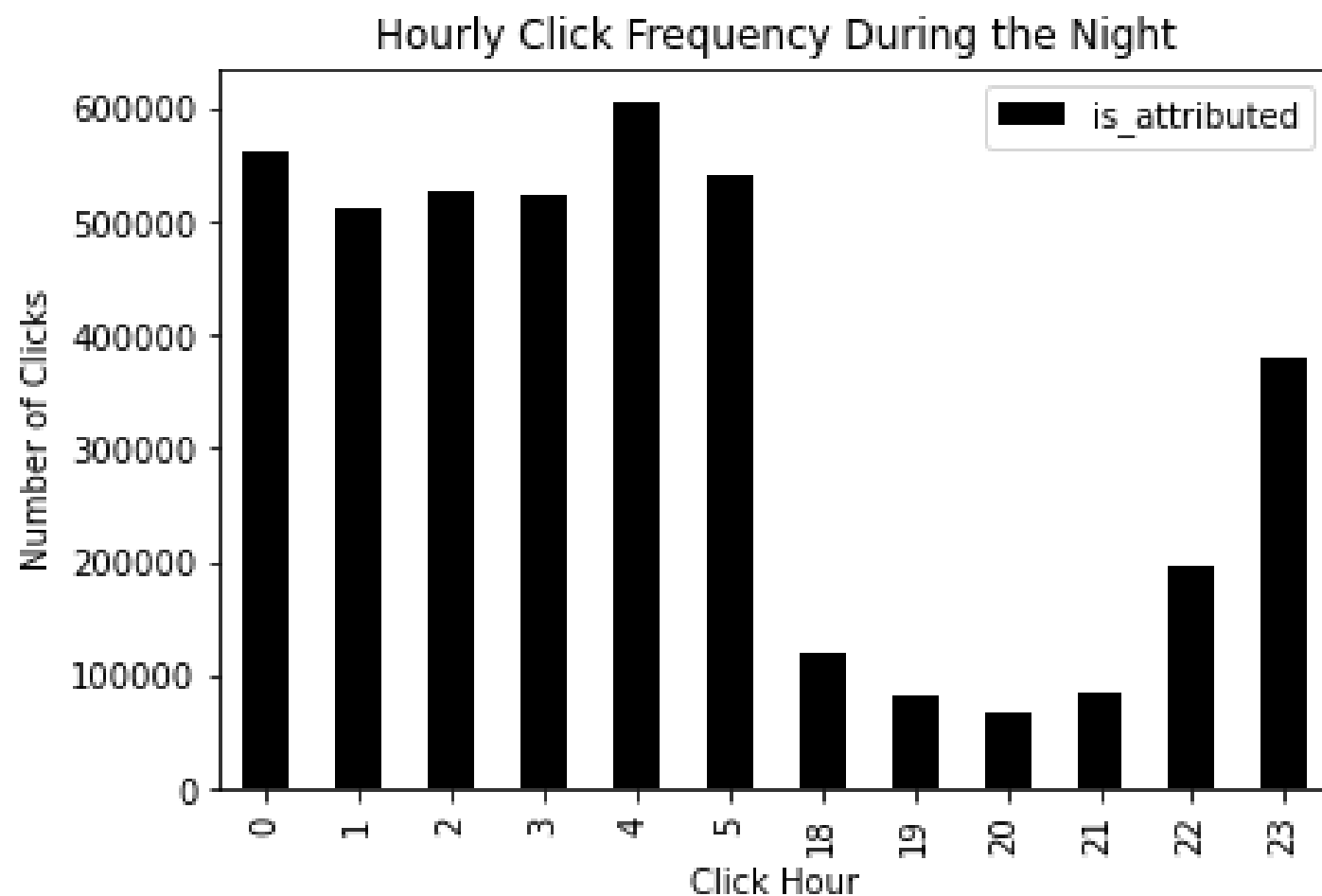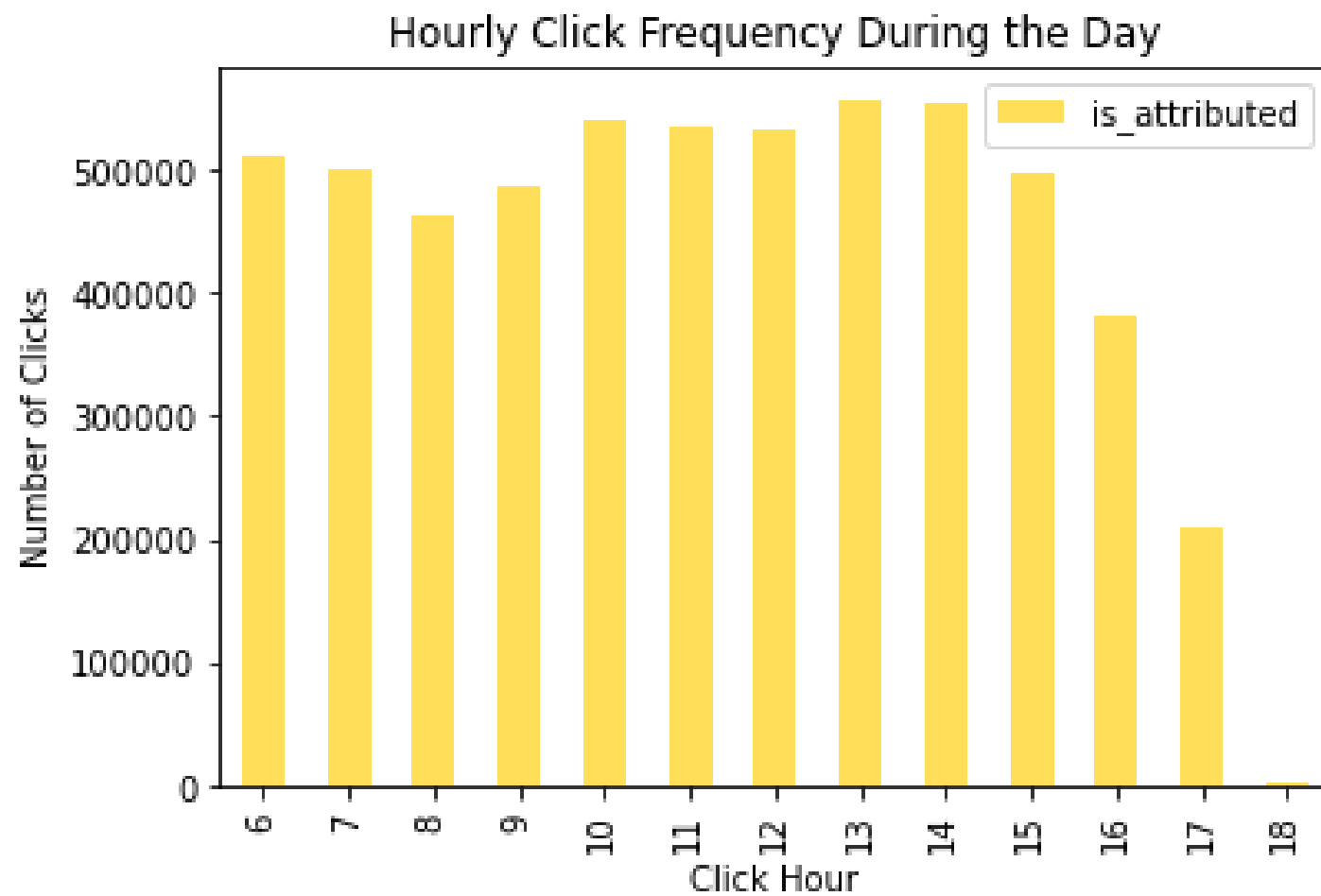

App Downloaded vs Not Downloaded

05

HOURLY CLICK FOR FRAUD

Day Clicks: 6am - 6pm

Night Clicks: 6pm - 6am

Hourly Click Frequency During the Day

Hourly Click Frequency During the Night

x̄ of click count during the day

443,844.54

s of click count during the day

161,953.74

x̄ of click count at night

350,434.33

s of click count during the night

220,021.98

t-test

P-value=0.24

Welch's t-test

The test gives a t-statistic of 1.2 and a p-value of 0.24

07

CONCLUSION

Hypothesis results

Fail to Reject:

Since our P-Value Is greater than our alpha of 0.05 we can fail to reject our null hypothesis testing that the rate of fraud at night equals the rate of fraud during the day.

08

Next Steps

Main takeaways and how can we
use our data

Fraud occurs 24/7, It's best to not waste resources
targeting time of fraud occurance.

Dig deeper into non fraudulent data, take a look at
the hours that clicks actually do lead to people
downloading the app.

Create digital marketing campaigns to target
audiences that download the app at peak hours.

THANK YOU!!!

https://github.com/priyaseejoor

https://www.linkedin.com/in/priyaseejoor/

CONTACT