1. Good afternoon everyone.

My name is Kin-Yip, and I’m going to introduce you to the Chicago parking tickets data set as well as explore some revenue development strategies using this data.

2. The city of Chicago has a long history of public corruption dating back over 150 years.

A 2019 report by the Illinois Policy Institute estimated that public corruption cases cost the state of Illinois 550 million dollars every year between 2000 and 2017.

In order to maintain its reputation as the most corrupt city in America, the City of Chicago must find ways to unlock the full potential of its revenue streams.

3. I originally hoped to follow in the footsteps of ProPublica, the investigative journalism organization that compiled this data set, and examine how parking fines burdened different segments of Chicago society, but I was not able to find demographic data from the Census Bureau website in a format that would be easy to merge with my data set.

So I rescoped and determined that I could use the data I had to investigate the levers that affect ticket revenue.

4. The data set I am working with was compiled by ProPublica from FOIA requests to the City of Chicago for parking ticket records.

The data span January 1, 1996 to May 14, 2018.

Every record corresponds to a ticket issued.

In total, there were over 54 million records, with 22 features from the City of Chicago and a further 14 features engineered by ProPublica

Features contained information about the date and location a ticket was issued, the violation, information about the vehicle and zipcode it was registered to, the issuing officer, fine amounts, and the ticket status as of May 14, 2018 (was it paid, unpaid, dismissed, etcetera.)

5. The uncompressed csv is nearly 20 gb and too large to load into memory. I used unix to systematically sample every 25th record in order to produce a data subset with 2 million records that could be loaded into python.

From there, I used pandas to explore the data.

6. Now I’m going to show you some plots that demonstrate the information that can be gleaned from this data set.

Here is a plot of the number of tickets issued by the City of Chicago aggregated by year, with another superimposed plot of the fine amounts paid by year. You can see that although the number of tickets issued has declined since 2002, the revenue drawn by the City has been robust to that decline.

8. This is a plot of the disposition of various categories of tickets as of May 14, 2018. I categorized the tickets as paid, unpaid, or dismissed. I need to look into it so I won’t speculate as to the differences in distributions between violations.

10. One question I was interested in was whether raising the fine on a violation affected the probability of the ticket being paid.

Here I have plotted the cumulative proportion of tickets paid in blue, and the price of a ticket in green. Because there are so many observations, it is not easy to see if the slope of the blue line changes when fines increase.

11. So I turned to hypothesis testing. Here I performed a difference in proportions hypothesis test. I took a window of tickets issued in the 6 months before and after a fine increase as my two samples and compared the proportion of tickets paid. A simplifying assumption is that the two samples are independent, but I would need to spend more time in my data checking that assumption, as some drivers receive multiple violations and make up a nontrivial proportion of the total tickets issued, so they may have been ticketed both before and after a fine increase.

The difference in proportions test is a parametric test that relies on the Central Limit Theorem, which applies to my data as the windows capture a few hundred observations.

Here I have sketched the calculations involved in the difference in proportions test.

12. Level 1 fines are issued for a violation, and level 2 fines can be assessed for unpaid tickets and enduring violations.

In the period of my data, fines for rush hour parking violations were increased three times. I windowed tickets 6 months before and after the fine increase and present here the p-values associated with one-sided difference in proportions tests.

There is not enough evidence to reject the null hypothesis that raising ticket fines decreases payment probability.

15. In conclusion, the City of Chicago should raise fines to increase revenue, as drivers don’t seem to be sensitive to fine increases.

16. Generate a data set of change in payment probabilities and change in fine amount (e.g. (0.65 – 0.70, $200 - $100) => (-0.05, 200)

(1 - 0.65/0.70, $200/$100) => (0.07, 2)

Investigate the correlation between change in payment probability and change in price.

Investigate other violations.

Investigate revenue amounts rather than probability of payment, as increasing fines may discourage drivers from offending in the first place.

17. And these are my references. Any questions?