Capstone III: Analysis of Affordable Rentals in US from 2010-2016

Tam Nguyen 04.2023

Agenda

- The dataset (slide 3)

- The research questions, hypotheses and how will data answer those questions (slide 4)

- The hypotheses (slide 5)

- How will the findings be used? (slide 6)

- The analyses - question #1 (slide 7)

- The analyses - question #2 (slide 8, 9)

- The analyses - question #3 (slide 10, 11)

- Insights and conclusions (slide 12)

The dataset

The dataset being used contains rental affordability records for different regions in the U.S. For reference, the original dataset is [here](#).
There are 1713 records across nine fields, but there are six fields being used in this analysis as follows:
• RegionName
• Tier
• Date
• Income
• Zri
• PercentIncomeSpentOnRent

The Zillow Rent Index (ZRI) is the median Rent Zestimate® (estimated monthly rental price) for a given geographic area on a given day, and includes the value of all single-family residences, condominiums, cooperatives and apartments in Zillow's database, regardless of whether they are currently listed for rent. It is expressed in dollars.

The research questions, hypotheses and how will data answer those questions

1.   Is there any differences percentages spent on rent in different tiers and in different states?

2. How the Zri (Zillow rental index) trend in New York and California in difference tiers from 2010-2016?

3. Between New York and San Francisco, how different the percentage spent on rent in different tiers: any significant difference of percentages spent on rent in upper tier or bottom tier between those cities?

- Visualizations (scatter plots) will show a correlation for percentage spent on rent in different states and in different tiers

- Visualizations (scatter plots) will show a correlation for Zri in different states and in different tiers

- Z-test will show a significant different of average percentage spent on rent between New York and San Francisco on the upper tier and bottom tier.

The hypotheses

There are two hypotheses related to the previous research questions (question #3) as follow:

- My first hypothesis is that there is no difference between the percentage spent on rent of the upper tier in New York and San Francisco.
- My second hypothesis is that there is no difference between the percentage spent on rent of the bottom tiers in New York and San Francisco.

I will test this at 95% significant level

How will the findings be used?

- The people looking for rent will be interested in knowing which region has the percentage spent on rent that fit with their budget.

- Investors looking for opportunities investing in rent can also look for the rental index trends for each area. It can be helpful to target for their invest.
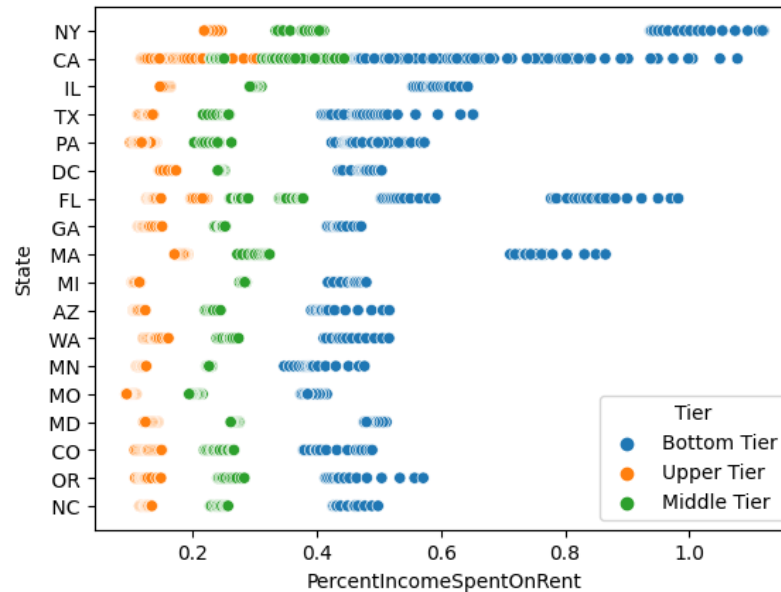
The analyses - question #1

For question #1: Is there any different percentages spent on rent in different tiers and in different states?
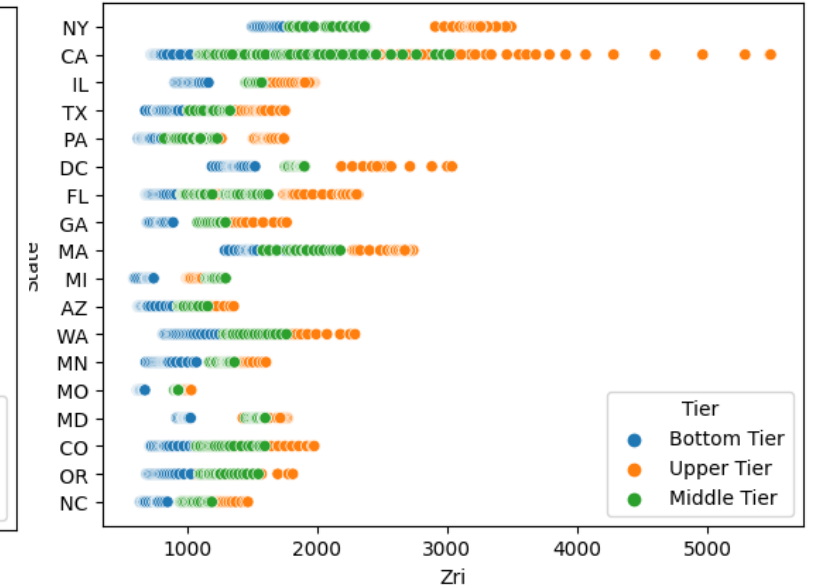Visualizations (scatter plots) will show a correlation for percentage spent on rent in different states and in different tiers, also Zri index.
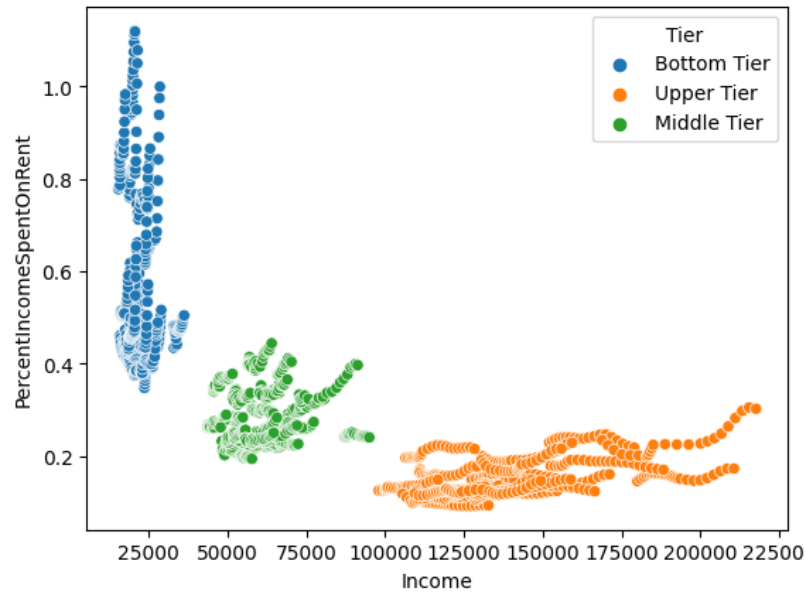
Bottom tier has a wide range from 0.3-1.1 percent income spent on rent. The middle tier is 0.2-0.4 and the upper tier has a very consistent percentage spent on rent around 0.2%

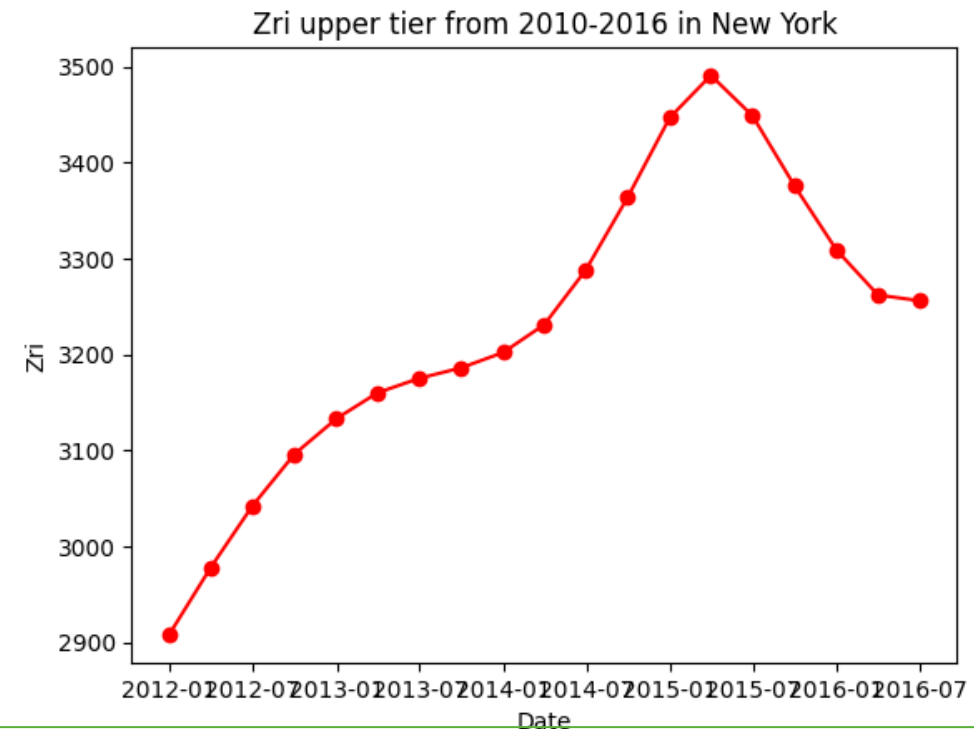**NY and CA has the highest percentage income spent on rent in the bottom tier**
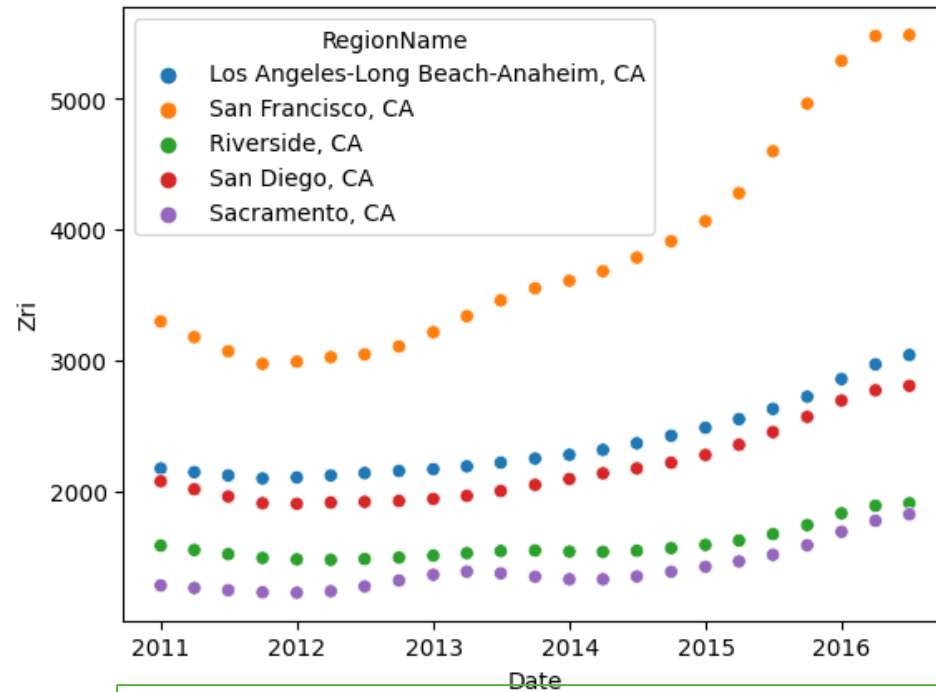
NY and CA has the highest Zri index

The analyses - question #2

How the Zri index trend in New York and California in different tiers from 2010-2016?

Visualizations (scatter plots) will show the Zri index trend in the upper tier in California and NewYork

- **In CA, Zri index in the upper tier increases from 2010-2016, especially in San Francisco, but it is steady in the LA area, RiverSide, San Diego and Sacramento**

- **In NY, it was increasing until 2015 and cooling down after that**

The analyses - question #2 continued:

How the Zri trend in the bottom tier in New York and California from 2010-2016?
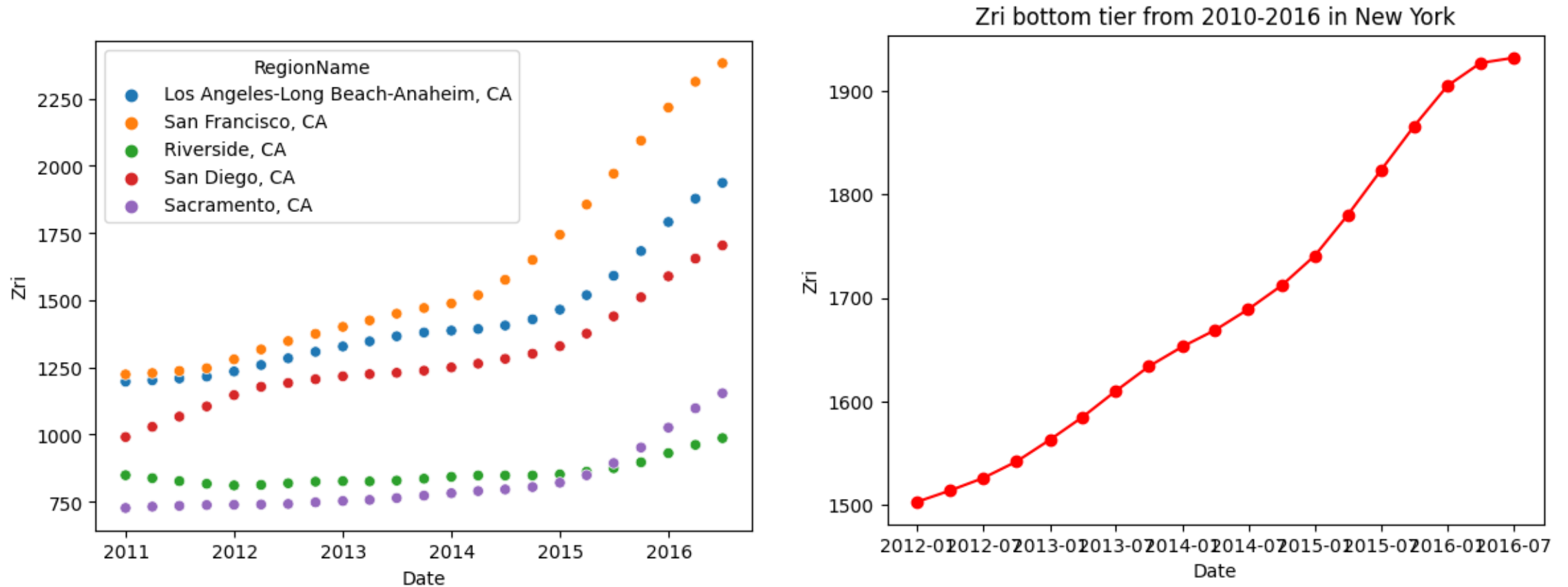
Zri index in the bottom tier increases from 2010-2016, especially in San Francisco and New York

The analyses - question #3

For the first hypothesis:

Null hypothesis: there is no difference between the percentage spent on rent of the upper tier in New York and San Francisco.

Alternative hypothesis: there is a difference between the percentage spent on rent of the upper tier in New York and San Francisco.

An independent samples z-test was used to support the hypothesis:

```
 (z=-0.17282606760930402, p=0.8627881463665664)

P>0.05: fail to reject the null
```

➡ **There is no difference between the percentage spent on rent in the upper tier in New York and San Francisco**

The analyses - question #3 continued:

For the second hypothesis:

Null hypothesis: there is no difference between the percentage spent on rent of the bottom tier in New York and San Francisco.

Alternative hypothesis: there is a difference between the percentage spent on rent of the bottom tier in New York and San Francisco.

An independent samples z-test was used to support the hypothesis:

```
(z=-9.44154986567535, p=3.673056082354075e-21)
```

```
P<0.05: Reject the null hypothesis
```

➡️ **There is a significant difference between the percentage spent on rent in the bottom tier in New York and San Francisco**

Insights and conclusions

- In the US from 2010-2016, the bottom tier income has a wide range from 0.3-1.1 percent income spent on rent. The middle tier is 0.2-0.4 and the upper tier has a very consistent percentage spent on rent around 0.2%. New York and California has the highest percentage income spent on rent in the bottom and upper tiers

- In CA, rent (Zri index) in the upper tier increases strongly in San Francisco, but it is steady in the LA area, RiverSide, San Diego and Sacramento. In New York, it was increasing until 2015 and cooling down after that. While Zri index in the bottom tier grows strongly in those cities from 2010-2016

- Confirm hypothesis #1: There is no significant difference between the percentage spent on rent in the upper tier in New York and San Francisco

- Reject hypothesis #2: There is a significant difference between the percentage spent on rent in the bottom tier in New York and San Francisco

For reference:
The proposal is here
The Colab Notebook is here