Chloe
Ha
Gabe
Finn
Pierce

1. Look at the data, and come up with questions that the data can help you to answer.

We're finding more trends

Research Questions

- What are the contrasts between tech, law, and finance lease details?
    - E.g. more commonly high/low quality, big or small, number of offices, etc?
    - Compare leases and card swipes?
- How much data do we have (proportion) for tech, law, and finance?
    - We have about half as much data for law as we do for tech and finance. About 20k for law, then 40k for tech, finance.
- Generally, where are each of these sectors trying to get lease locations? Has it changed time?
- Who are the most common companies getting leases? (the biggest client)
    - Wework has the most common leases at 197, followed by Knotel (82), Regus (76), Industrious (62), and Amazon (58).
- What are the characteristics of these leases for each company?
- Are they moving to the suburbs over time, as opposed to Central Business Districts?
    - How does this compare between regions and industries?
        - CBD is preferred across the board, but much more so for law than finance or tech.
        - Across all 3 industries, suburbs have become more common over time since 2022.
- Does unemployment affect leasing rates (based on location or industry)?
    -
- Are the square footage of the company and the company's size correlated?
- How does the square footage of tech, law, and finance change over time?
    - Square footage of each lease, on average, has been increasing after the COVID-19 pandemic. This happens across tech, finance, and law, with the finance industry needing significantly more space than tech and law firms.
- Do higher-occupancy markets also show higher lease volume?
    - Manhattan shows high occupancy and high lease volume
    - Austin and SF have moderate occupancy but lower leasing activity
    - No strong positive correlation across most markets => companies may still be hesitant to make long-term leasing decisions
- Which sectors are paying the highest rents in 2023-2024? How has the changed over time?

    - Finance pay the highest rent on average, followed by tech and laws ⇒ reason: they value location prestige
    - We know that we now prefer hybrid work trends but these industries are still committing to expensive space
- Do regions with higher unemployment have less leasing activity or smaller leases?
    - We often assume that higher unemployment is a red flag for businesses ⇒ but in this chart,
    - On overall, South region had the highest leasing with moderate unemployment ⇒ strong demand and market confidence
    - Northeast and West showed positive relationship
    - Midwest had lower lease counts on overall, with
- Big General questions I came up with: How have commercial leasing patterns changed across since the pandemic, and what should companies consider when choosing where and how to lease office space today?

Covid

- What are the trends before and after the COVID pandemic?
    - 2018-2019
    - 2022-2023
    - 2023-2024
- Using secondary data source: Which industries are returning to the office the fastest?

Sector

- What regions have the most common property leases for tech, law, and finance? Does this change over time?
    - Northeast always have the highest amount of leases, peaked in 2022
    - Financial services always have a highest amount of lease
    - **Upward trend across all regions and industries from 2023-2024 -> potential increase prediction?**
- Have the companies shifted their leasing strategy (such as smaller leases size)?
    - Definitely a shift but not significantly differ by industries
    - Potentially have to clean data due to missing values in the 'leasing' features
    - Also there are trends but they got cut off in 2022 so no potential
- What types of companies are leasing in most markets?

Region

- How have the leases changed in San Francisco from 2018 during pre-covid and after covid?
    - Number of Leases dropped for all industries during covid and are now on the rise towards or at pre-covid levels.
- How has office space utilization demand changed post-COVID by region? (direct_available_space) + sublet increase/decrease (suggest underutilized space)
    - Market with high available space for rent: Manhattan, Downtown Chicago, Chicago suburb, Houston, Atlanta, Boston, Los Angeles

- Are companies moving away from high-cost cities (SF, Manhattan) to more affordable cities (Atlanta, or Austin)? -> leasing trend
    - Average price

Rent cost

- Employment rate relative to rent price (how price-sensitive our clients are)
- What should customers know about: Which markets currently offer high value (low rent, high availability, high demand)

Potential Ways to Organize our Research

- Section: Law, Tech, Finance
- Location:
    - General OR specific
    - Northeast (NY, Pennsylvania)
    - South (Atlanta)
    - Midwest/Central (chicago)
    - West (california-like)
    - **Challenge:** very big regions
    - **Challenge:** consolidate data from city-specific data and general data
- Before/After COVID
- Specific trends: are we looking for changes over LOCATION, changes over TIME, changes over size/traits of leases that people are asking for?

Helpful Code

Combining unemployment with leases

```
# want a single row per year, quarter, and state. We want the AVERAGE
unemployment for it - it was the unemployment before.
new_unemployed = unemployment_df.groupby(['year', 'quarter',
'state']).mean().reset_index()

def label_unemployment(row):
    year = row.year
    quarter = row.quarter
    state = row.state

    # .loc finds a row, the column search makes a series. We want to only
get the value. We turn it into a list
    return new_unemployed.loc[((new_unemployed['year'] == year) &
(new_unemployed['quarter'] == quarter) & (new_unemployed['state'] ==
state))]['unemployment_rate'].tolist()[0]


leases_df['Unemployment Rate'] = leases_df.apply(label_unemployment,
axis=1)
```

Cleaning unnecessary columns

```
leases_df = leases_df.drop(columns=['building_name', 'building_id',
'address', 'zip', 'costarID'])
'''Probably not helpful columns'''
leases_df = leases_df.drop(columns=['internal_submarket'])
```

Separate by quarter

```
leases_df["quarter_numeric"] = leases_df["year"] + (0.25 *
((leases_df["quarter"].str.slice(1).astype(int)) - 1))
```

Pandas filter based on finance, tech, or law industries

```
leases_finance_df = leases_df[leases_df['internal_industry'] == 'Financial Services and Insurance']

leases_tech_df = leases_df[leases_df['internal_industry'] == 'Technology, Advertising, Media, and Information']

leases_law_df = leases_df[leases_df['internal_industry'] == 'Legal Services']
```

For Finn: change ‘market = “Chicago”’ for other cities

```
# Filter for only the three target industries

target_industries = [

    'Financial Services and Insurance',

    'Technology, Advertising, Media, and Information',

    'Legal Services'

]



market = "Chicago"



leases_filtered =
leases_df[leases_df['internal_industry'].isin(target_industries)]



leases_filtered = leases_filtered[leases_df['market'] == market]



# Group by year, region, and industry to count number of leases

firm_presence = leases_filtered.groupby(

    ['year', 'internal_industry', 'CBD_suburban']

).size().reset_index(name='firm_count')



# Plot
```

```
plt.figure(figsize=(16, 7))

sns.lineplot(

    data=firm_presence,

    x='year',

    y='firm_count',

    hue='internal_industry',

    style='CBD_suburban',

    markers=True,

    dashes=False

)



plt.title('Firm Presence by Region for Tech, Law, and Finance Over Time')

plt.xlabel('Year')

plt.ylabel('Number of Leases (Firm Presence)')

plt.legend(title='Market / Industry', bbox_to_anchor=(1.05, 1), loc='upper
left')

plt.tight_layout()

plt.show()
```

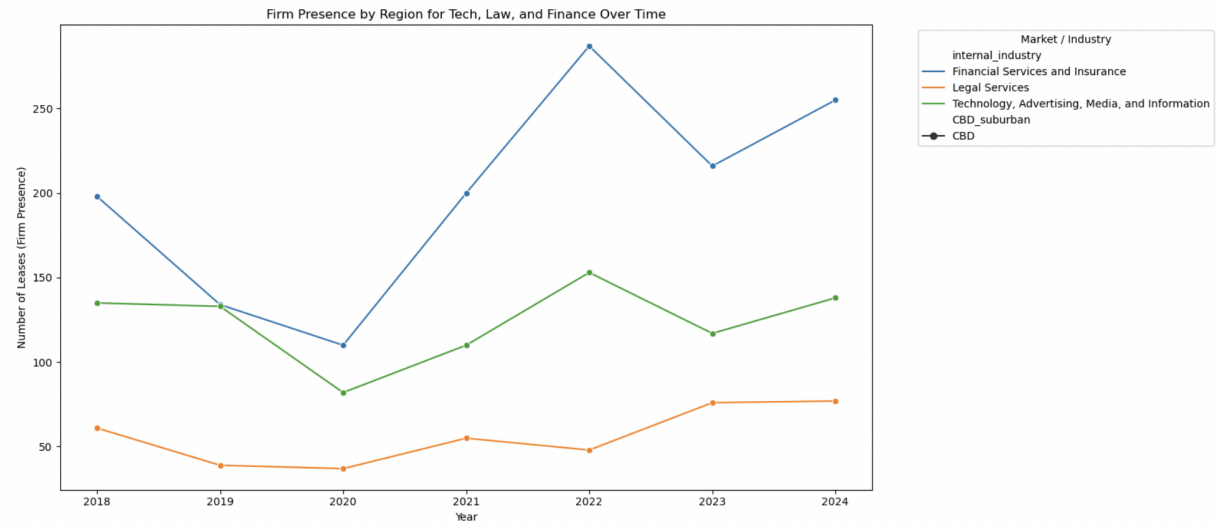
-

Los Angeles

- Notable Trends
    - Leases in all three industries and both markets increased from 2023 to 2024 indicating increasing demand
    - The CBD is preferred to Suburban markets with nearly triple the amount of leases across the three industries

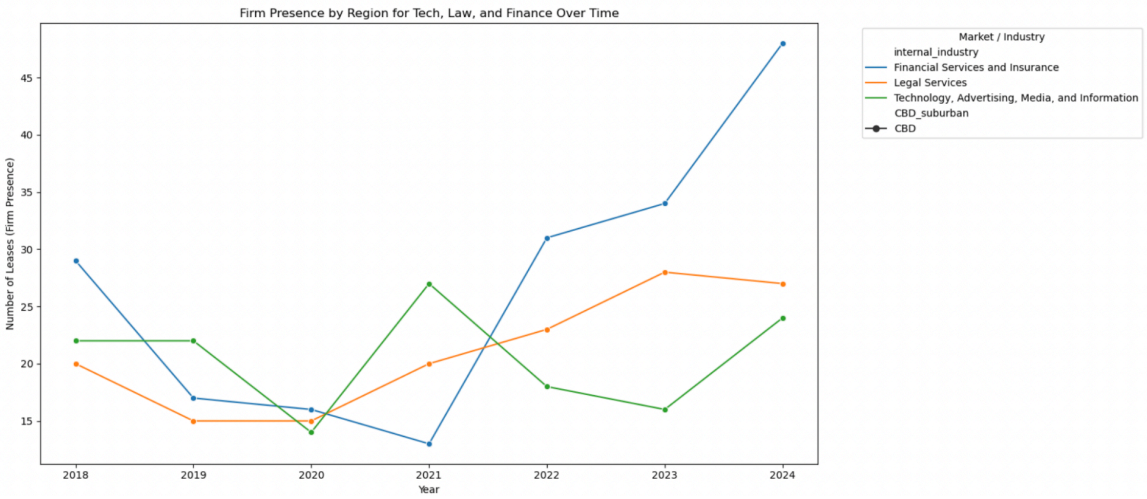
Boston

Manhattan



- Industries in Manhattan dropped to their lowest number of leases at the start of Covid but have since rebounded higher than their pre Covid numbers.
- No Suburban market

Chicago



- Notable Trends:
    -

- mplt.figure(figsize=(16, 7))

```
sns.lineplot(
    data=firm_presence,
    x='year',
    y='firm_count',
    hue='internal_industry',
    style='CBD_suburban',
    markers=True,
    dashes=False
)

plt.title('Firm Presence by Region for Tech, Law, and Finance Over Time')
plt.xlabel('Year')
plt.ylabel('Number of Leases (Firm Presence)')
plt.legend(title='Market / Industry', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()
```

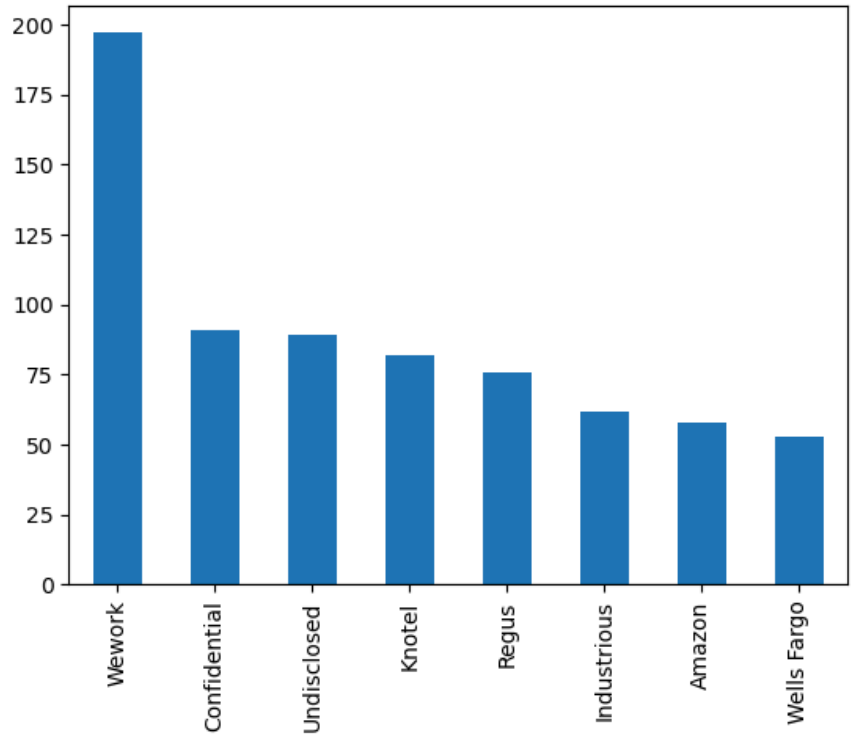
What regions have the most common properties leases for tech, law, and finance? Does this change over time?

```
# Filter for only the three target industries

target_industries = [

    'Financial Services and Insurance',

    'Technology, Advertising, Media, and Information',

    'Legal Services'

]

leases_filtered = leases_df[leases_df['internal_industry'].isin(target_industries)]



# Group by year, region, and industry to count number of leases

firm_presence = leases_filtered.groupby(

    ['year', 'region', 'internal_industry']

).size().reset_index(name='firm_count')



# Plot

plt.figure(figsize=(16, 7))

sns.lineplot(

    data=firm_presence,

    x='year',

    y='firm_count',
```

```
    hue='region',
    style='internal_industry',
    markers=True,
    dashes=False
)

plt.title('Firm Presence by Region for Tech, Law, and Finance Over Time')
plt.xlabel('Year')
plt.ylabel('Number of Leases (Firm Presence)')
plt.legend(title='Region / Industry', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()
—------
BY YEAR
# Filter for only the three target industries
target_industries = [
    'Financial Services and Insurance',
    'Technology, Advertising, Media, and Information',
    'Legal Services'
]

leases_filtered = leases_df[leases_df['internal_industry'].isin(target_industries)]

# Group by year, region, and industry to count number of leases
firm_presence = leases_filtered.groupby(
    ['year', 'region', 'internal_industry']
).size().reset_index(name='firm_count')

# Plot
plt.figure(figsize=(16, 7))
sns.lineplot(
    data=firm_presence,
```

```
    x='year',
    y='firm_count',
    hue='region',
    style='internal_industry',
    markers=True,
    dashes=False
)

plt.title('Firm Presence by Region for Tech, Law, and Finance Over Time')
plt.xlabel('Year')
plt.ylabel('Number of Leases (Firm Presence)')
plt.legend(title='Region / Industry', bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()

BY QUARTER
# Group by quarter
firm_presence_quarterly = leases_filtered.groupby(
    ['quarter_numeric', 'region', 'internal_industry']
).size().reset_index(name='firm_count')

# Plot with quarter-level detail
plt.figure(figsize=(16, 7))
sns.lineplot(
    data=firm_presence_quarterly,
    x='quarter_numeric',
    y='firm_count',
    hue='region',
    style='internal_industry',
    markers=False,
    dashes=True
)

plt.title('Quarterly Firm Presence by Region for Tech, Law, and Finance')
plt.xlabel('Time (Quarter)')
plt.ylabel('Number of Leases (Firm Presence)')
```

```
plt.legend(title='Region / Industry', bbox_to_anchor=(1.0
```

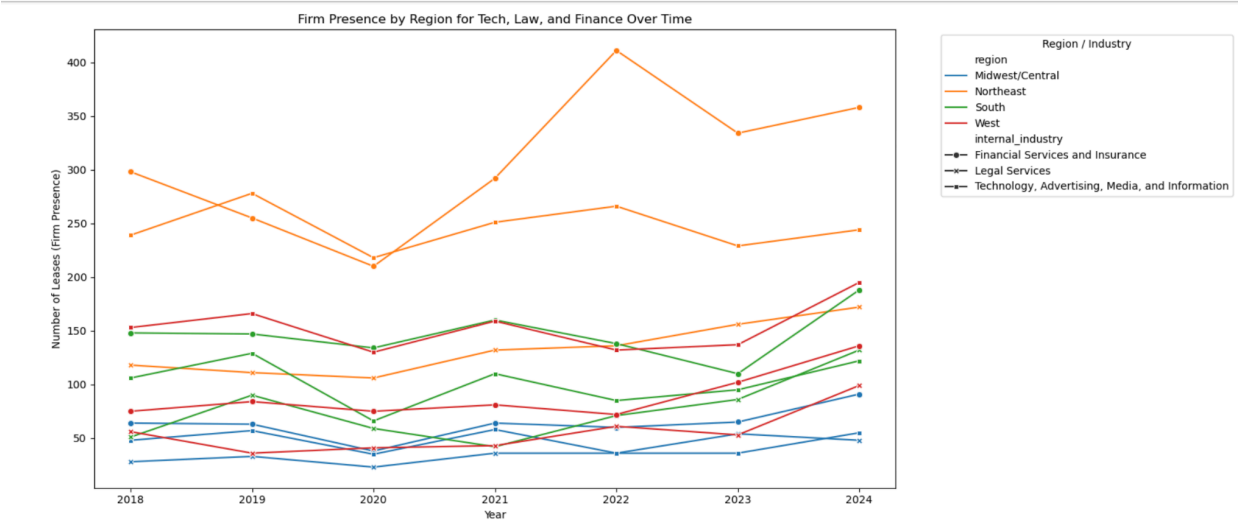



```
5, 1), loc='upper left')
plt.tight_layout()
plt.show()
```

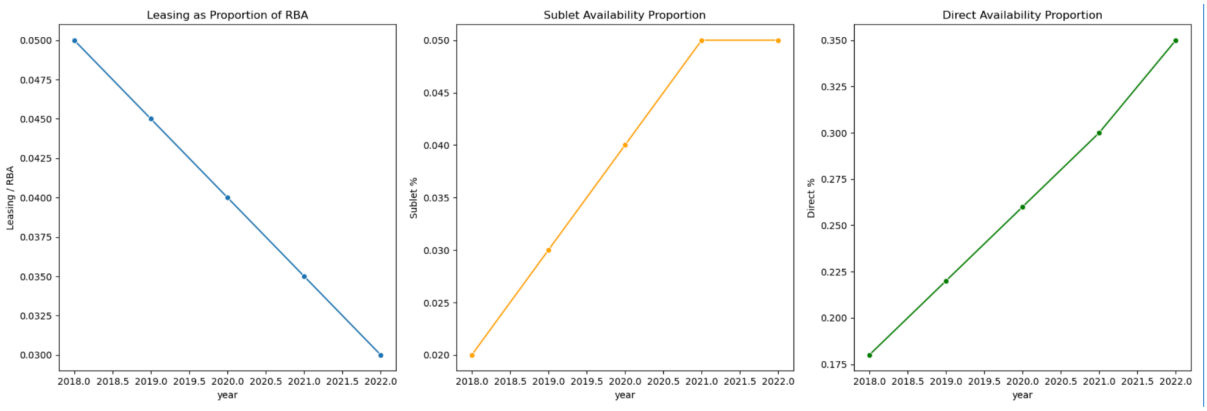
Visualizations

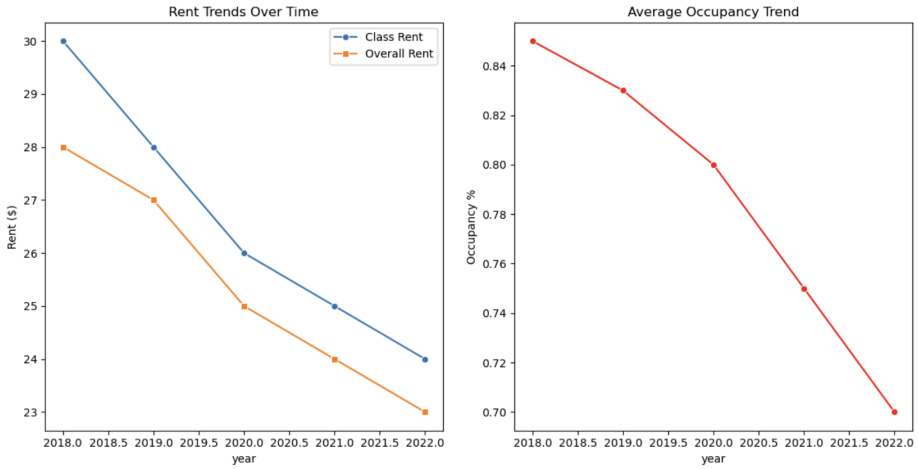
Number of companies with a certain lease

What regions have the most common property leases for tech, law, and finance? Does this change over time?



Have the companies shifted their leasing strategy (such as smaller leases size)?

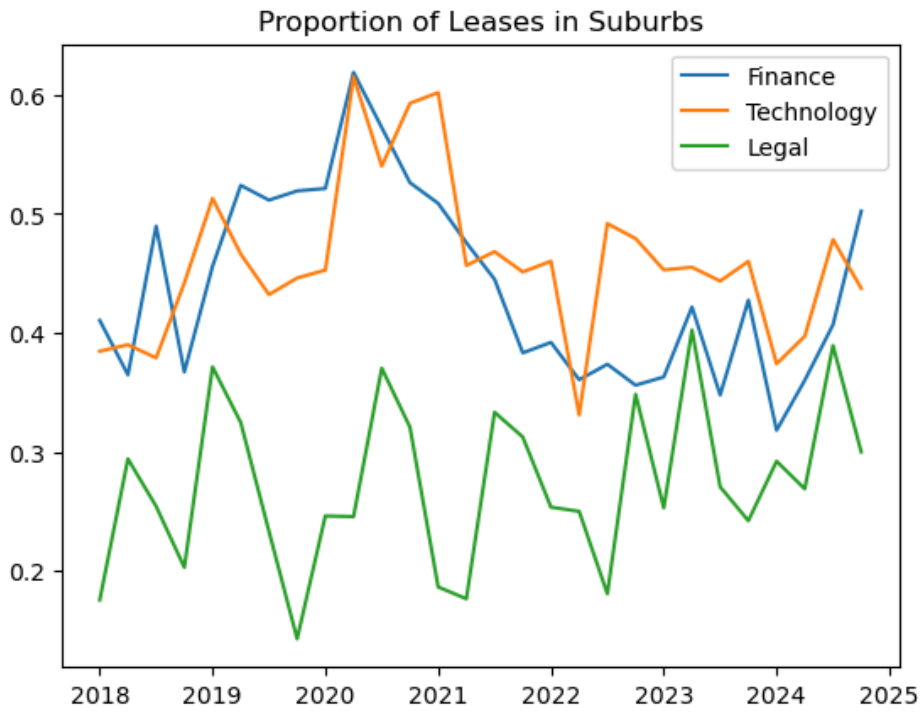




Average Lease Size Over Time by Industry

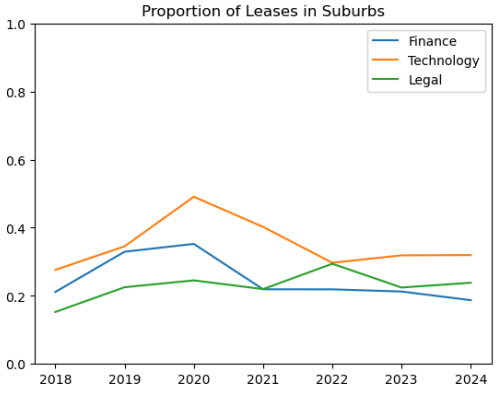
- Law and tech seem to have maintained remote work in some capacity post-pandemic, leading to less of a need for leasing extremely large areas.

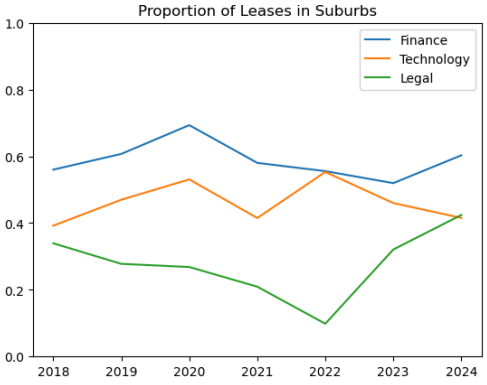
How have the proportions of leases over time changed based on industry?



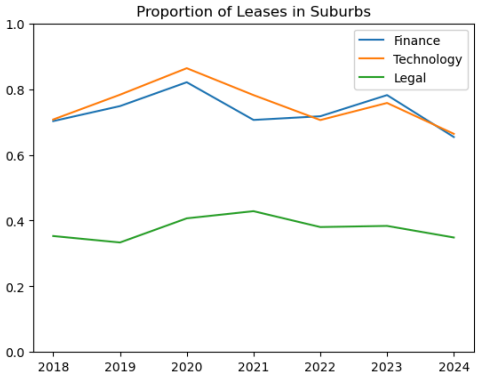
- Suburbs are more commonly leased in the south as opposed to northeast
- Legal tends to lease in CBD as opposed to finance and tech

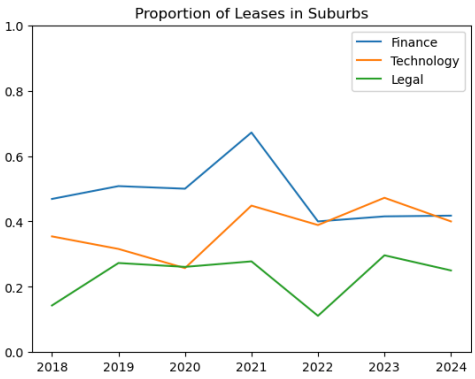
Northeast West





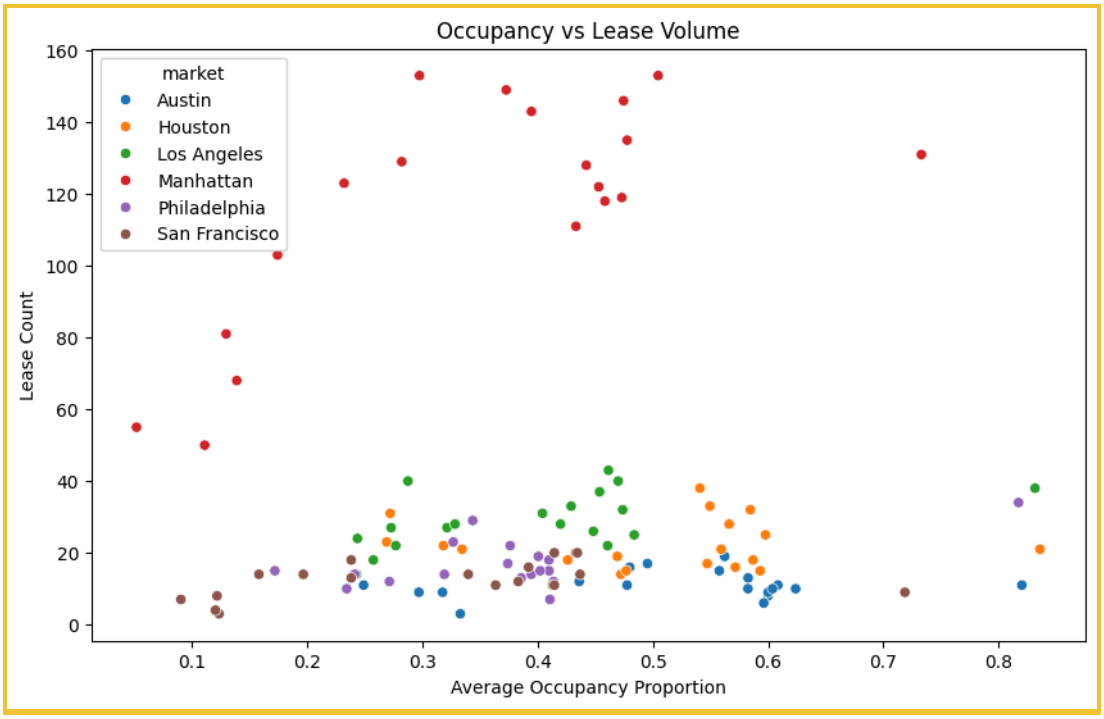
South Midwest





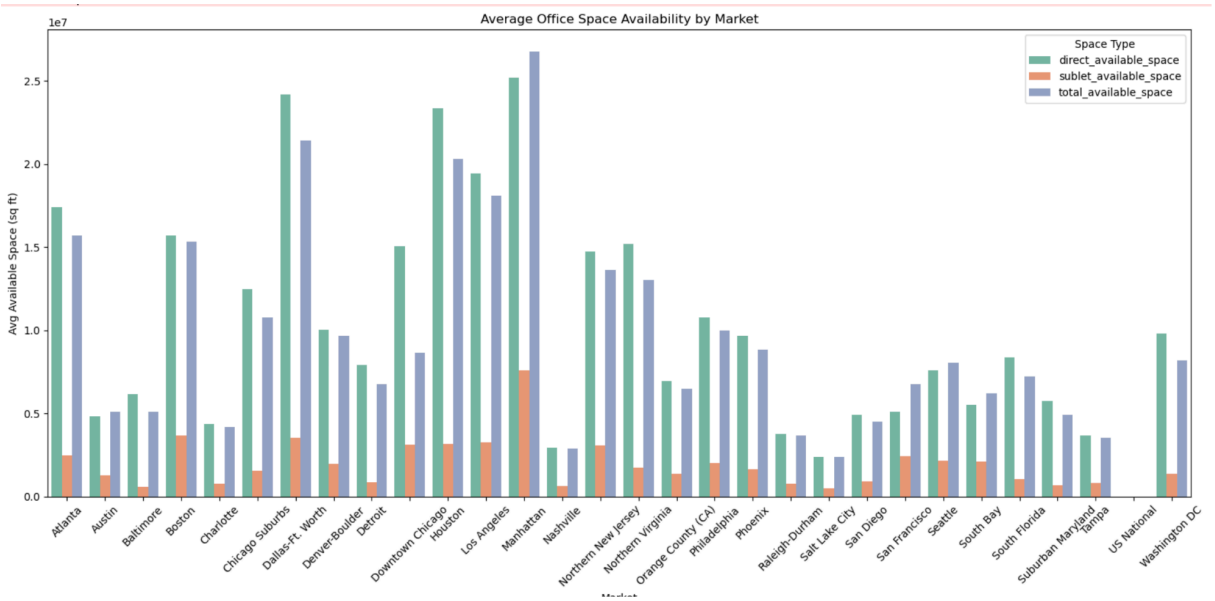
Do higher-occupancy markets also show higher lease volume?

Across the board, manhattan is higher. It goes down & is similar during covid when the lease counts are very low and the average occupancy proportion is low

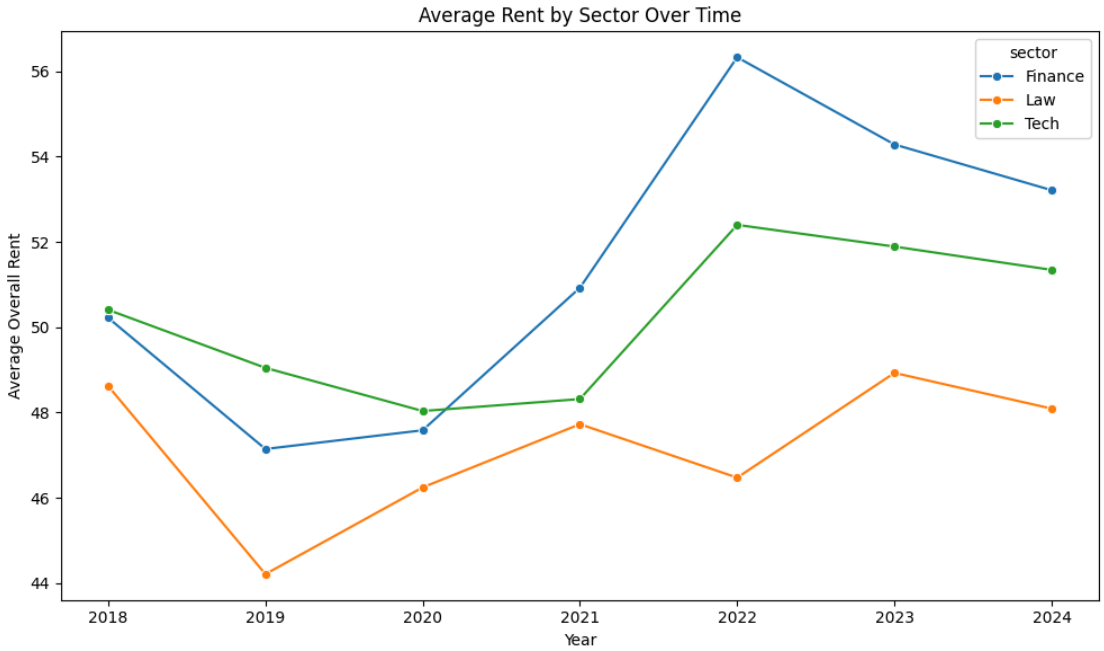


Average Office Space Availability by Market

Sort descending by city and then this can work to show “here’s how we can grow our work”

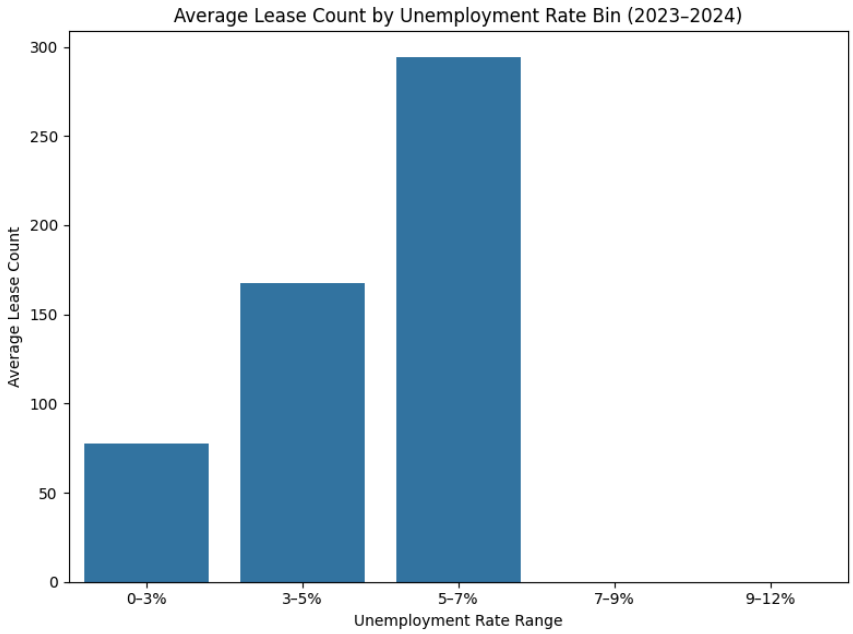


Which sectors are paying the highest rents? How has the changed over time?
Let's try rent per square feet to see if that gives any info.

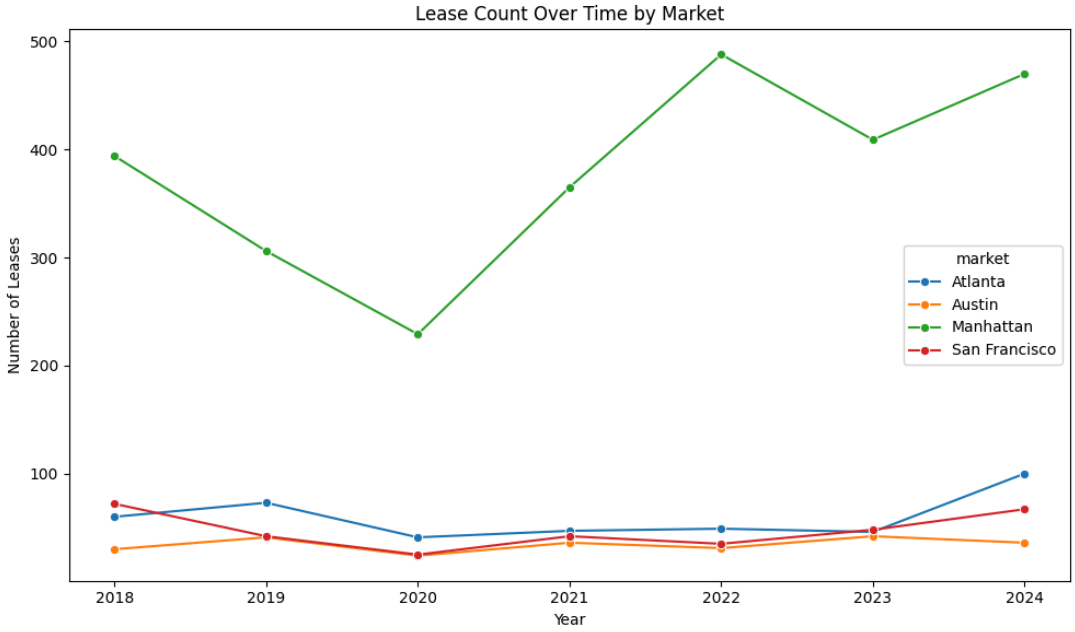


Which sectors are paying the highest rents? How has the changed over time?
Let's try rent per square feet to see if that gives any info.

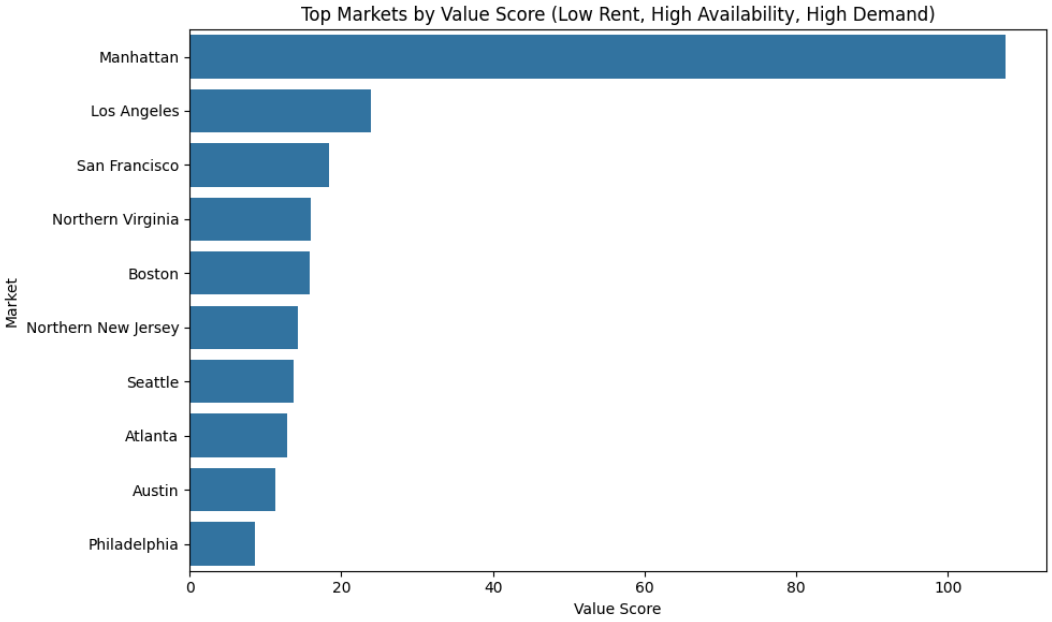
Do regions with higher unemployment have less leasing activity or smaller leases in 2023 - 2024?



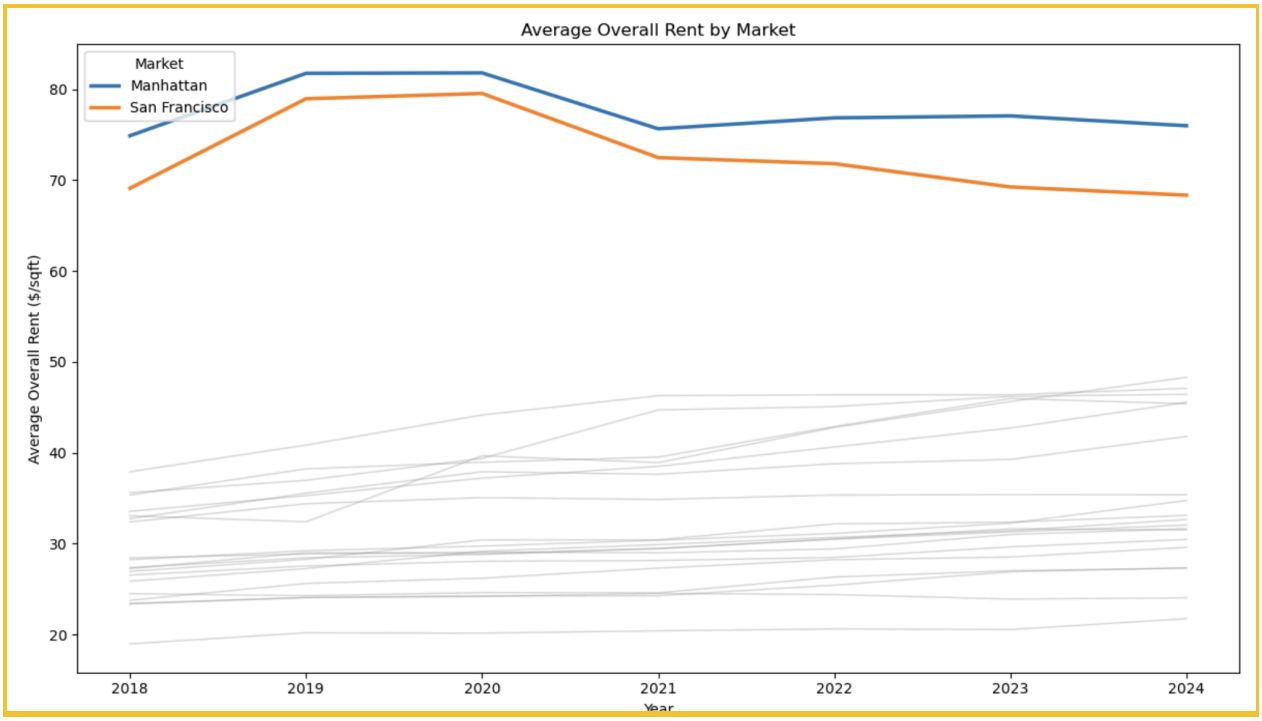
Are companies moving away from high-cost cities (SF, Manhattan) to more affordable cities (Atlanta, or Austin)? -> leasing trend



What should customers know about: Which markets currently offer high value (low rent, high availability, high demand)



Major Outliers here: Manhattan and San Francisco. Manhattan red, San Franciscos blue, everything else grey.



Helpful Terms in the Leases Dataset

- LeasedSF: leased square footage
- RBA (Rentable Building Area): the total square footage of a building that can be occupied or assigned to a tenant for the purpose of determining a tenant's total rental obligation.
- Internal Class: Either A or O. Different building designations. A indicates that a building is high quality and O indicates that it is not
- Available Space: The total amount of space that is currently being marketed for lease. It includes space that is vacant or also space that is currently occupied but will be
- vacant in the future. This is segmented into direct and sublet available space in the dataset as well
- Overall rent is in $/sqft, NOT $ overall. Good to know for past pierce so he didn't spend half an hour working with the wrong data.

For Pierce

We only have time to dive deeper into Manhattan, honestly.
So we have overall rent and available space for manhattan. We can find the rent/sqft over time, and also find whether its internal class is A grade or O grade. We want to predict, for a given lease, with 95% certainty what the rent should be per square footage for both A grade and O grade. (this is possible bc basically no suburb leases in Manhattan. it's all a big city). WRAP IT UP with the prediction

Also: does the rent/sqft increase as sq footage increases for manhattan? what about sanfrancisco (and its suburbs)? SANFRAN DOESNT HAVE SUBURBS.

Hypothesis: prolly costs more per square foot GENERALLY, but MORESO in CBD as opposed to in suburbs. So, suggestion would be to heavily consider leasing larger areas/recommending they move to cities with a bigger suburb presence, i.e. not Manhattan, not Chicago, etc.

This didn’t work out.