

# AdventureWorks2019 Analysis

---

**Group 3: Seb, Rory, Ahmed, Joanna,  
Abdirahman**

# Questions

- Q1) What are the regional sales in the best performing country?
- Q2) What is the relationship between annual leave taken and bonus?
- Q3) What is the relationship between Country and Revenue?
- Q4) What is the relationship between sick leave and Job Title (PersonType)?
- Q5) What is the relationship between store trading duration and revenue?
- Q6) What is the relationship between the size of the stores, number of

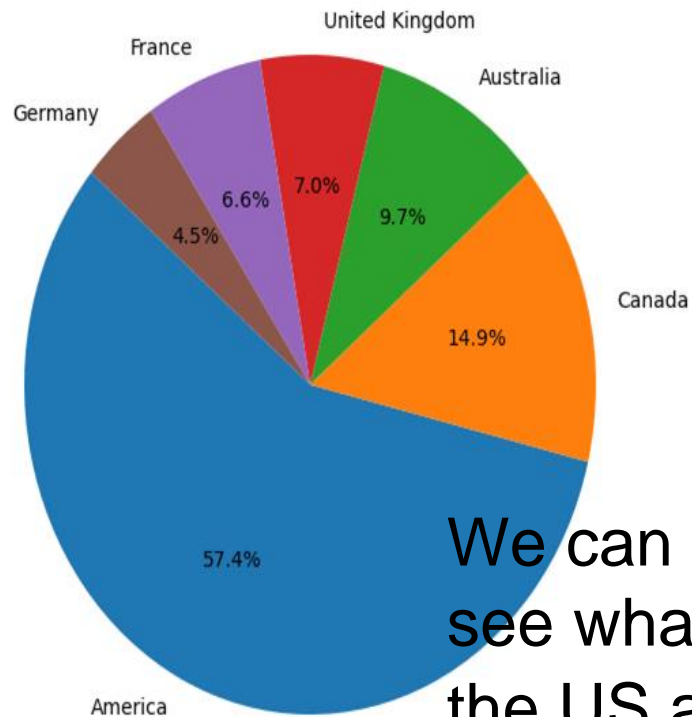
1. What are the regional sales in the best performing country?

---

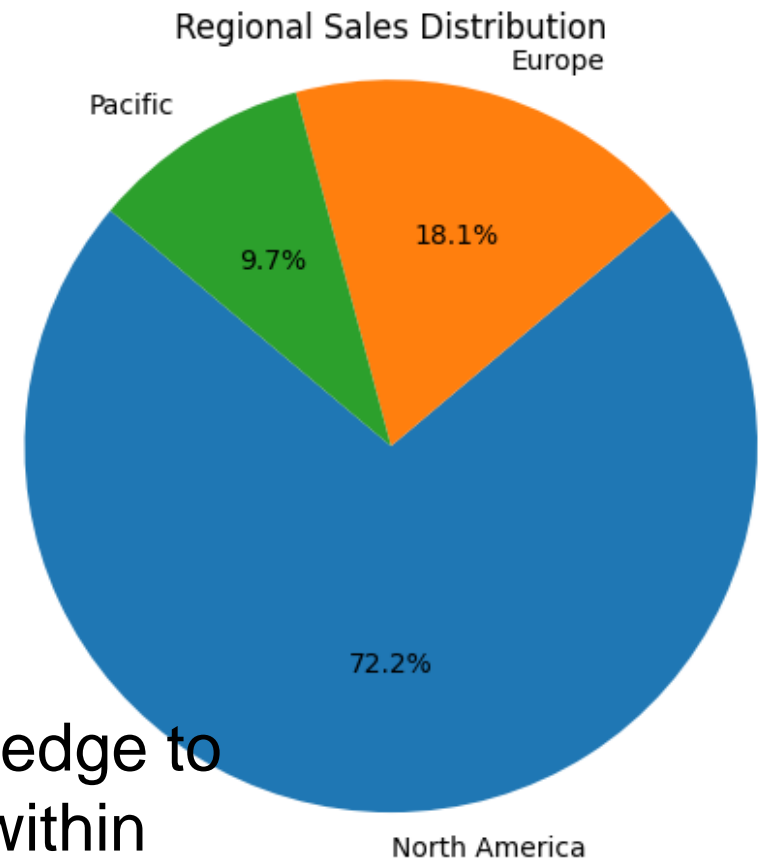
# The country with the greatest sales

---

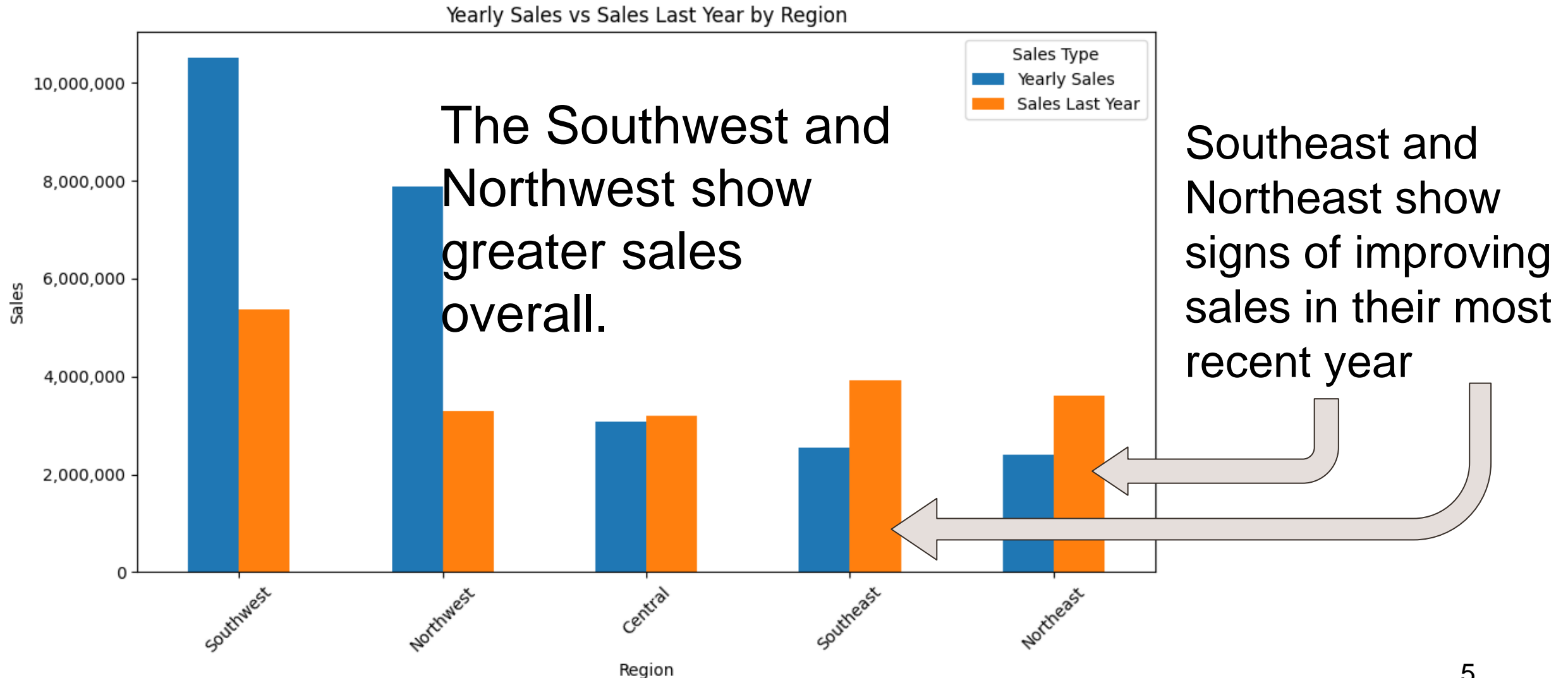
Firstly we can determine that North America performs the greatest overall



We can work from this knowledge to see what the regional sales within the US are



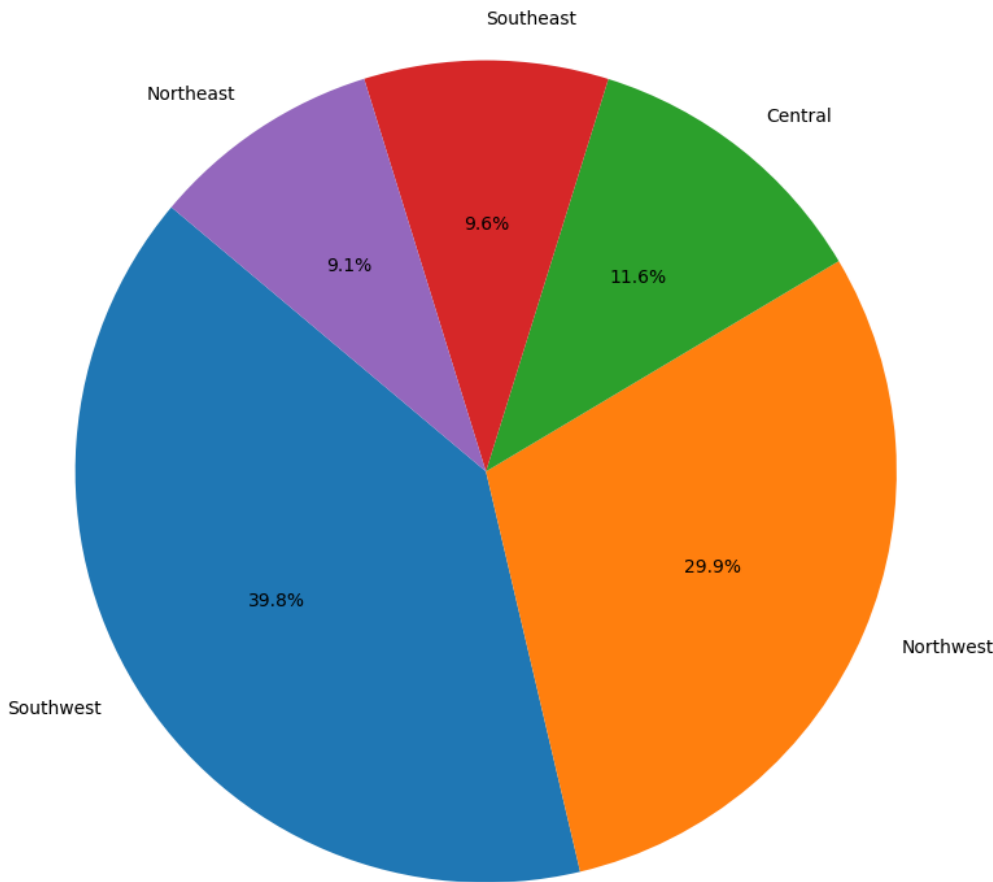
# Comparing regional sales in North America



# Looking at total regional sales to date

---

Yearly Sales Distribution by Region



If we look at the percentage sales distribution across the US regions we can see how the Southeast and Northeast outperform other regions by over **69%**

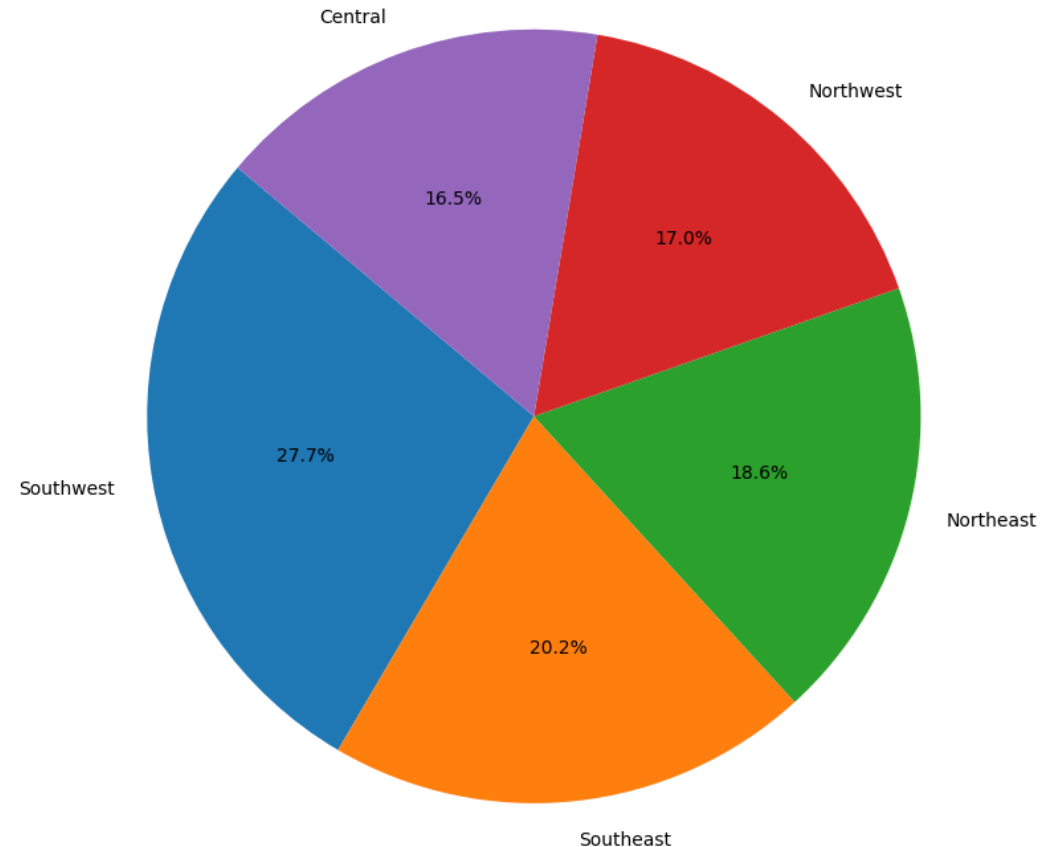
# Regional sales most recent performance

---

Looking at our most recent years sales we see much improved sales performance distribution

The Southeast has outperformed the Norwest

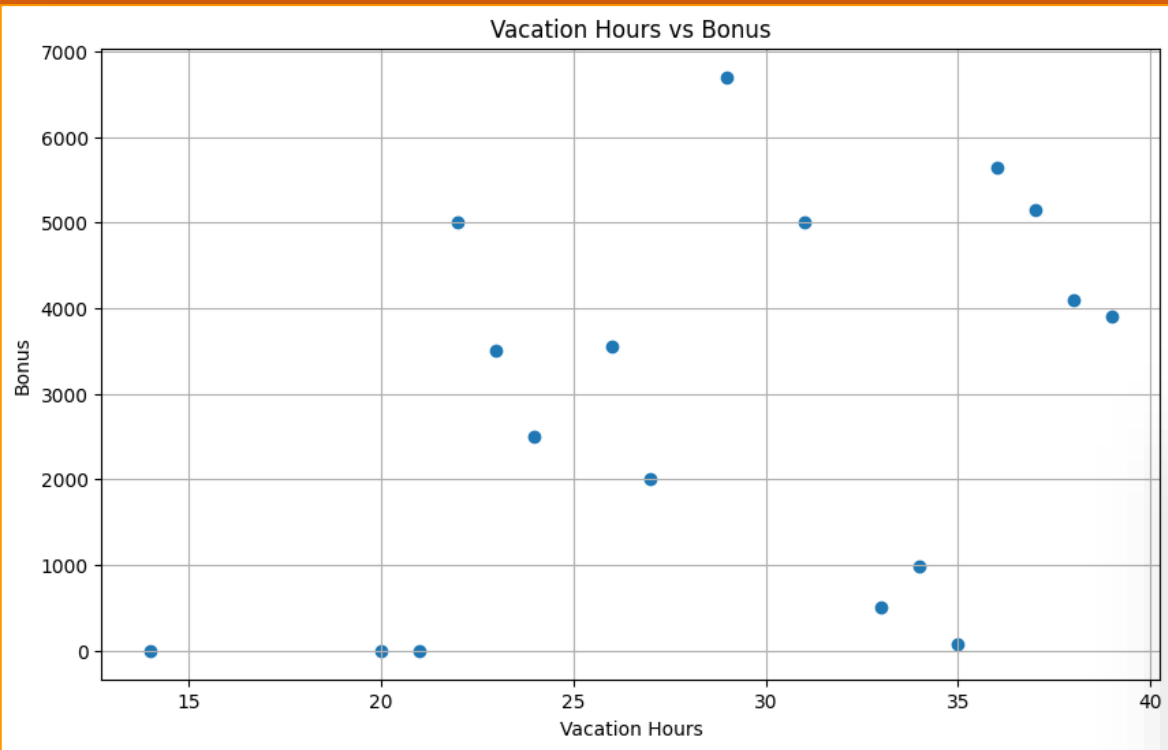
Yearly Sales Distribution by Region



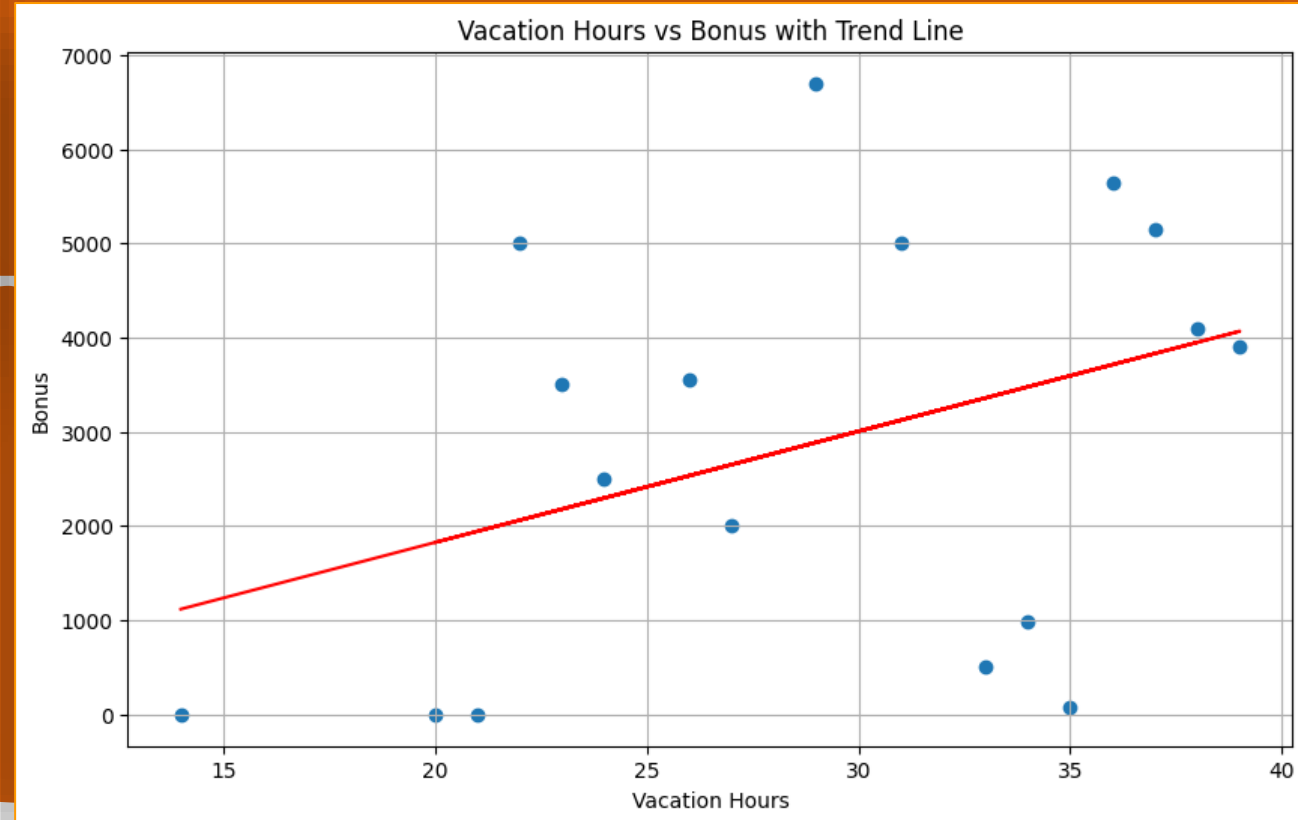
2. What is the relationship between annual leave taken and bonus?

---



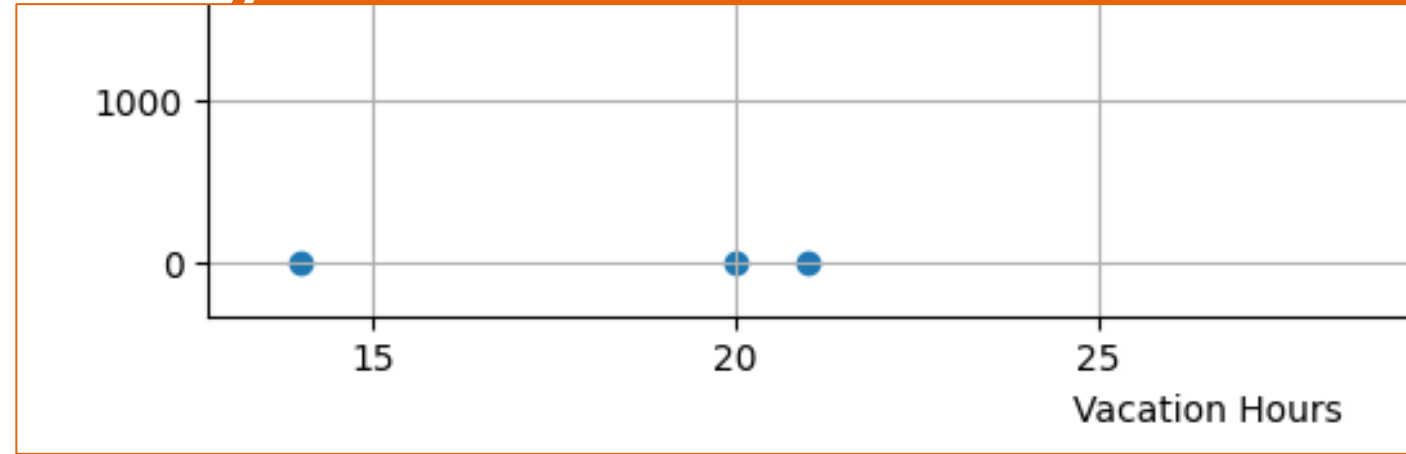


Looking at annual leave, our chart appears to show a counter-intuitive trend. The more annual leave, the more bonuses



Our trendline confirms this appears to be the case

These three zero values indicate employees who appear to be eligible for bonuses who do not get them.

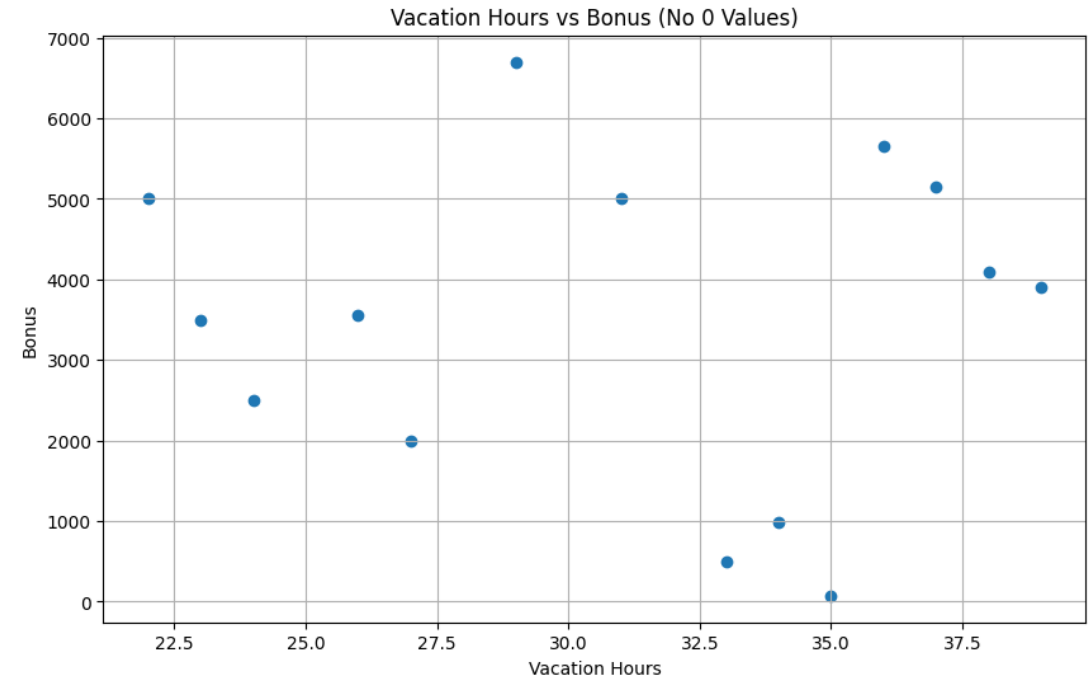
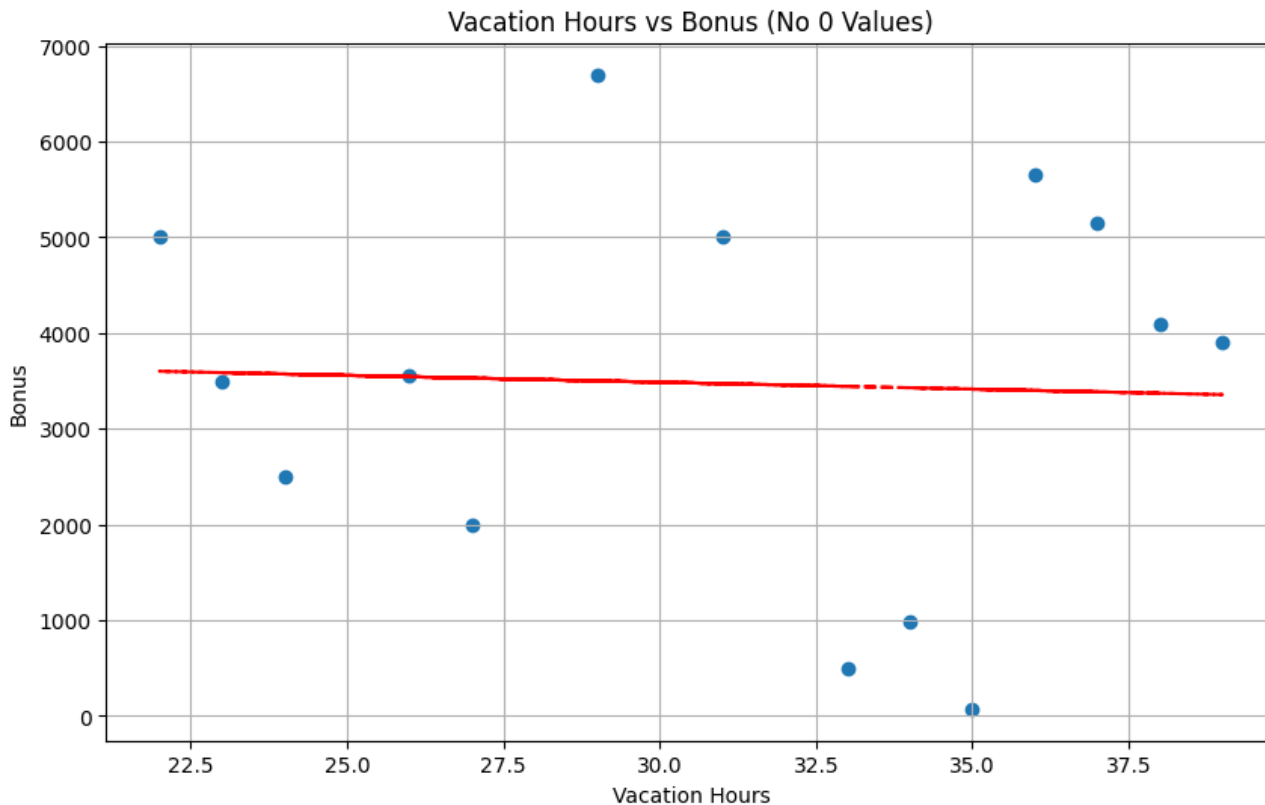


They also happen to be the ones who take the fewest vacation hours

They may be newer employees

Otherwise give them a raise maybe?

If we remove those those zero values as they do appear to be outliers we get a much more scattered scatter plot

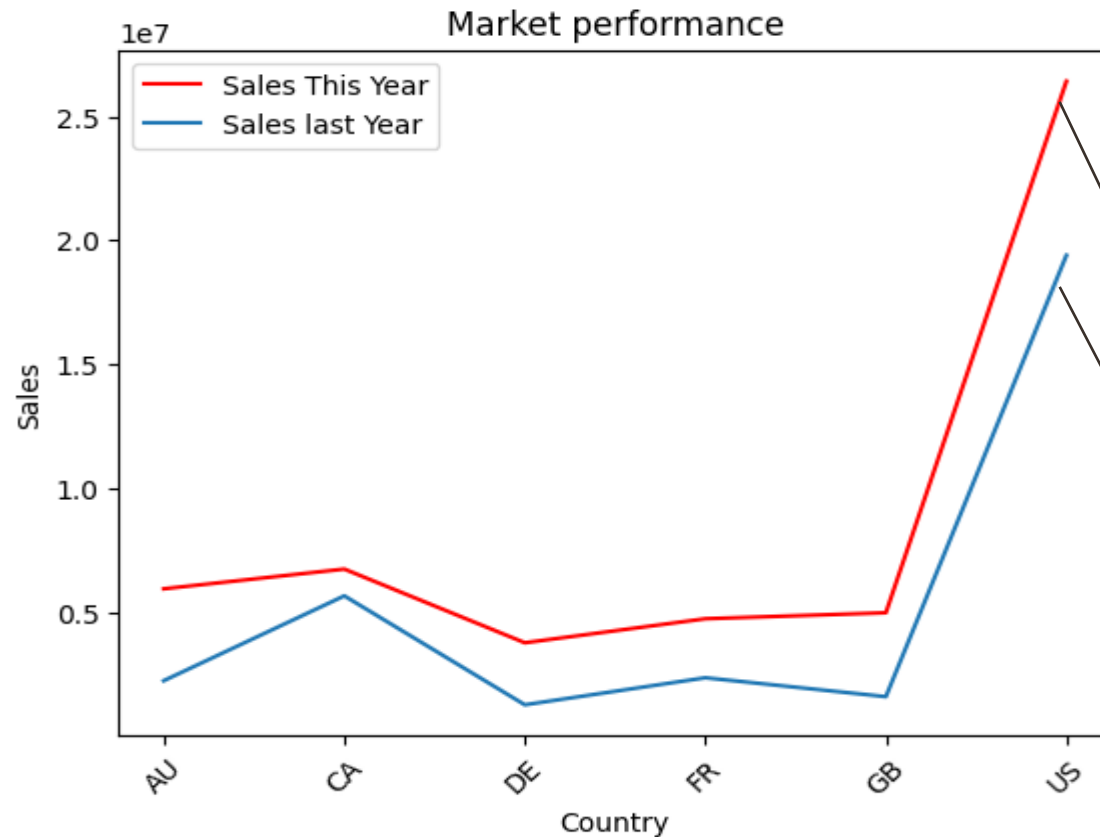


Our trendline confirms there is no real relationship. And in fact a slight negative correlation.

Q3) What is the relationship  
between Country and Revenue?

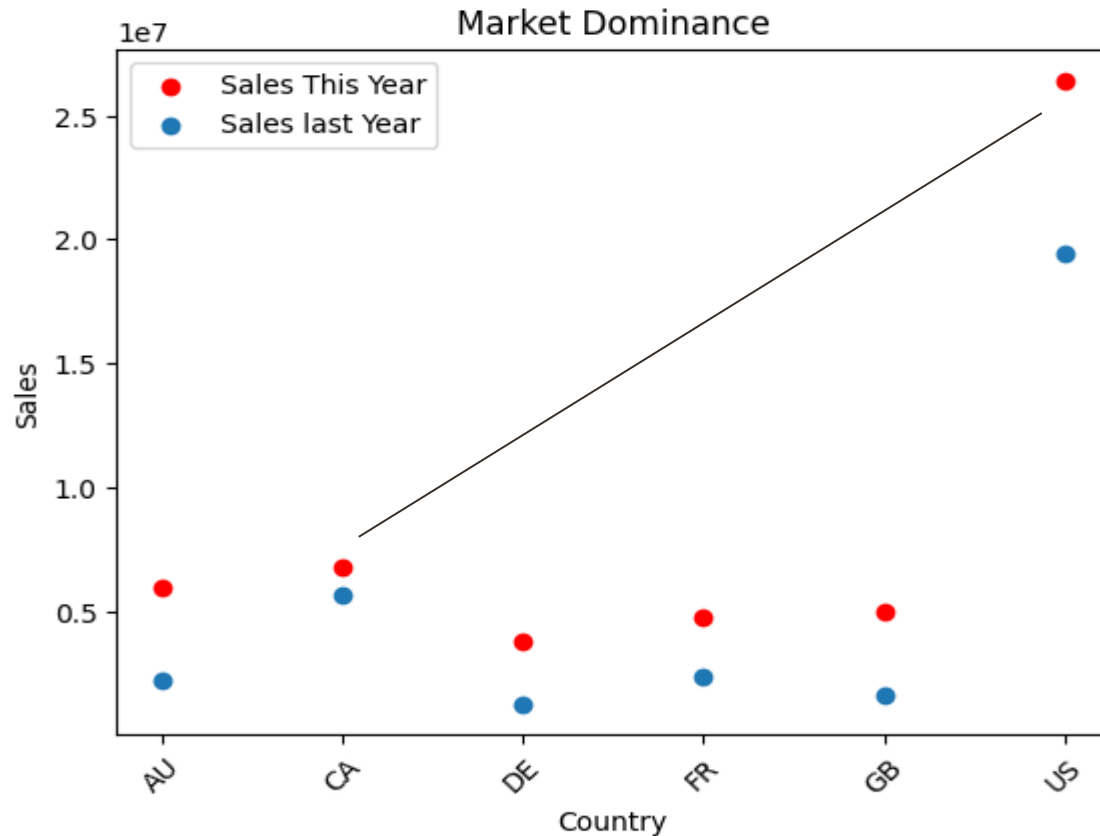
---

# Countries Performance



from the chart, we can see clearly that the US outperformed all the countries by selling  $2.641106e+07$  comparing to last year  $1.940250e+07$  the US sales increased by 36%.

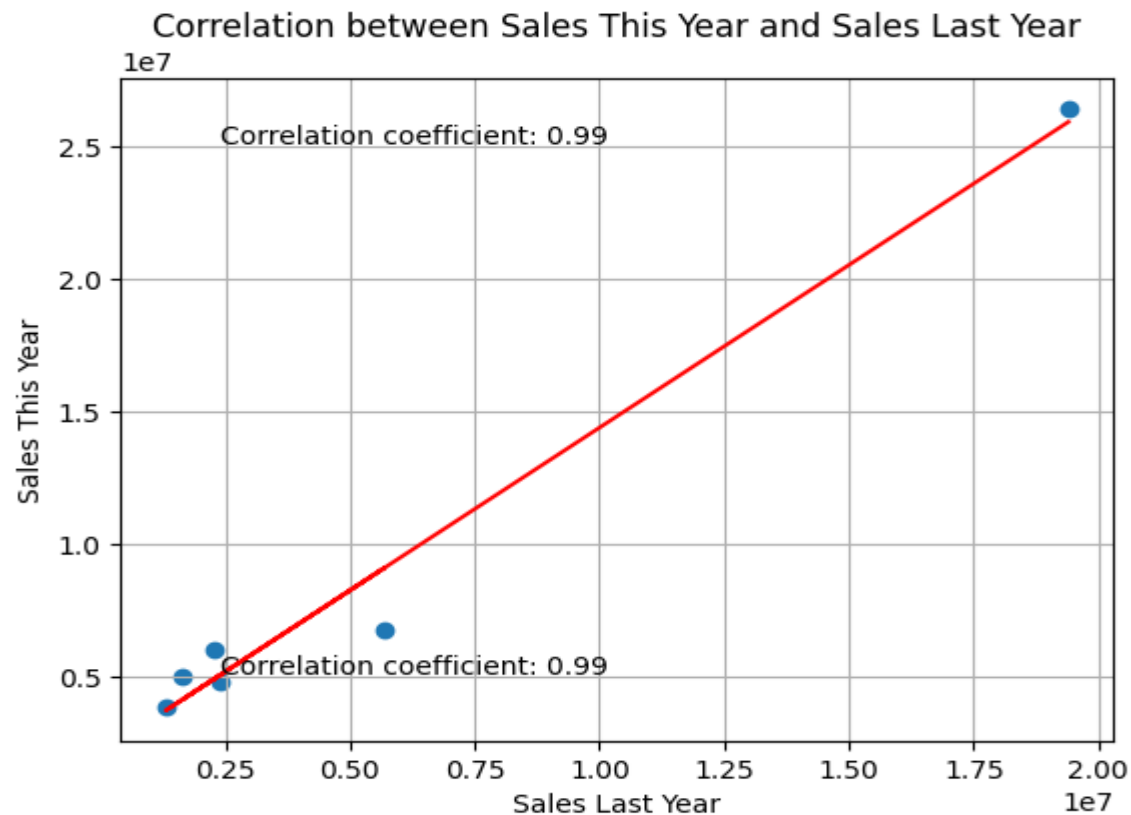
# Market Dominance



The chart illustrates the market dominance of the United States since last year, with a market share nearly three times compared to their competitor, which is Canada.

# Correlation between sales last and this year

---



This chart shows the correlation between countries.

The correlation is positive.

4. What is relationship between sick leave and Job Title (PersonType)?

---



# What is PersonType?

---

PersonType	Definition
SC	Store Contact
IN	Individual (retail) customer
SP	Sales person
EM	Employee (non-sales),
VC	Vendor contact
GC	General contact

- Not every person is eligible for SickLeaveHours
- For example:  
The company will not track a customers sick leave hours since they do not work for AdventureWorks

# Summary Statistics on Data

---

Cleaned data:

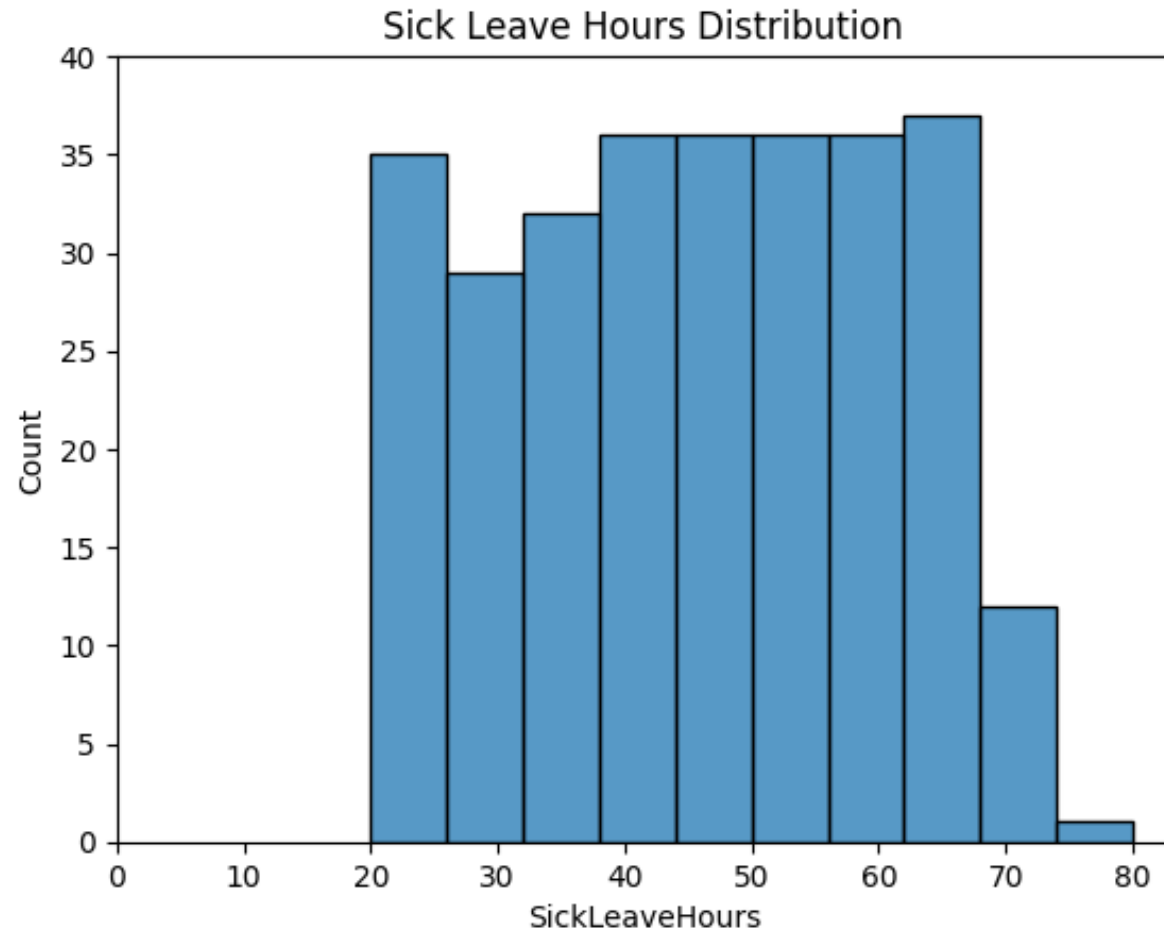
```
df_clean = df.dropna(subset=["SickLeaveHours"])  
print(df_clean.shape)
```

Calculate summary statistics:

	mean	std	sum	pct25	pct50	pct75
PersonType						
EM	46.003663	14.681796	12559.0	34.0	47.0	58.0
SP	34.117647	3.620814	580.0	31.0	34.0	37.0

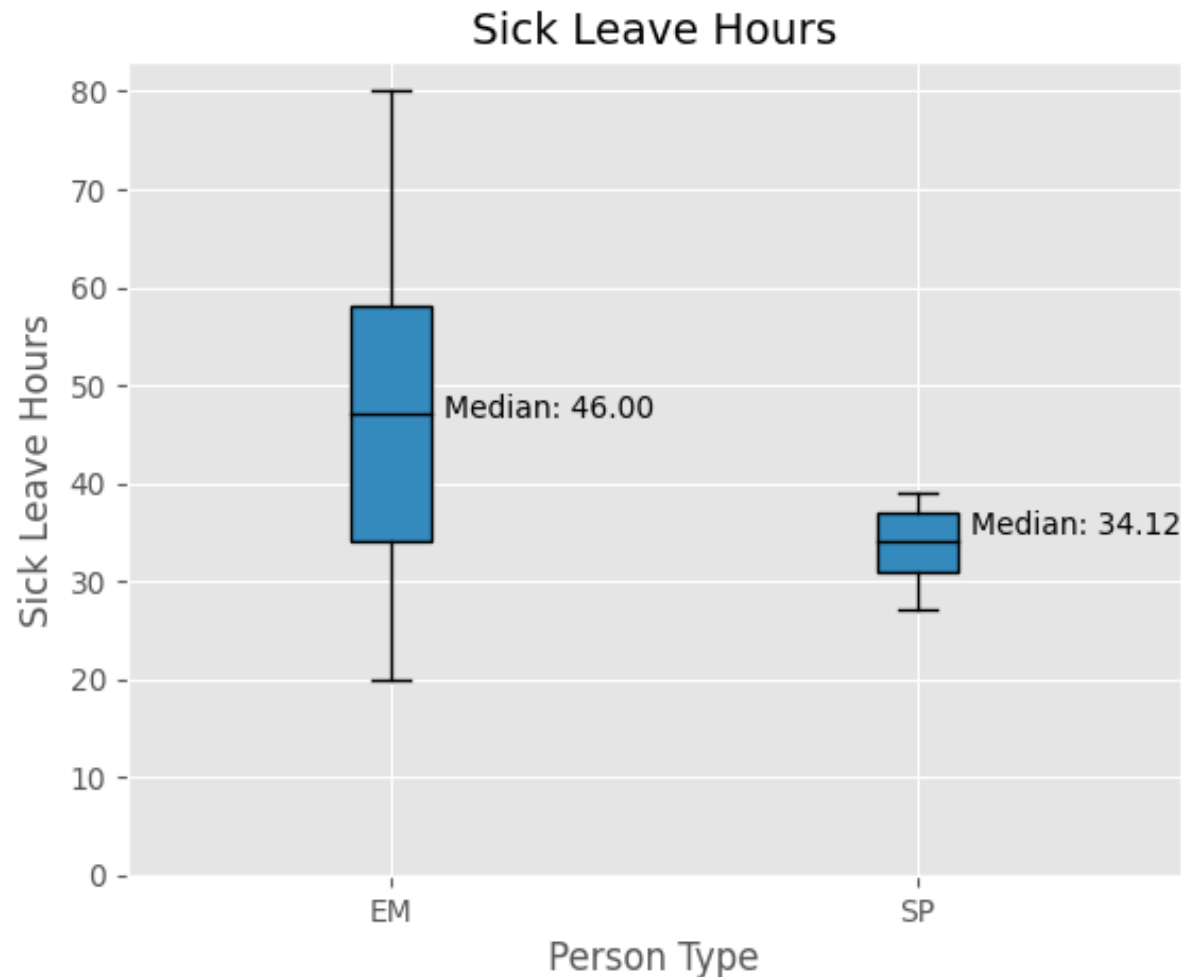
# Histogram of Sick Leave Hours

---



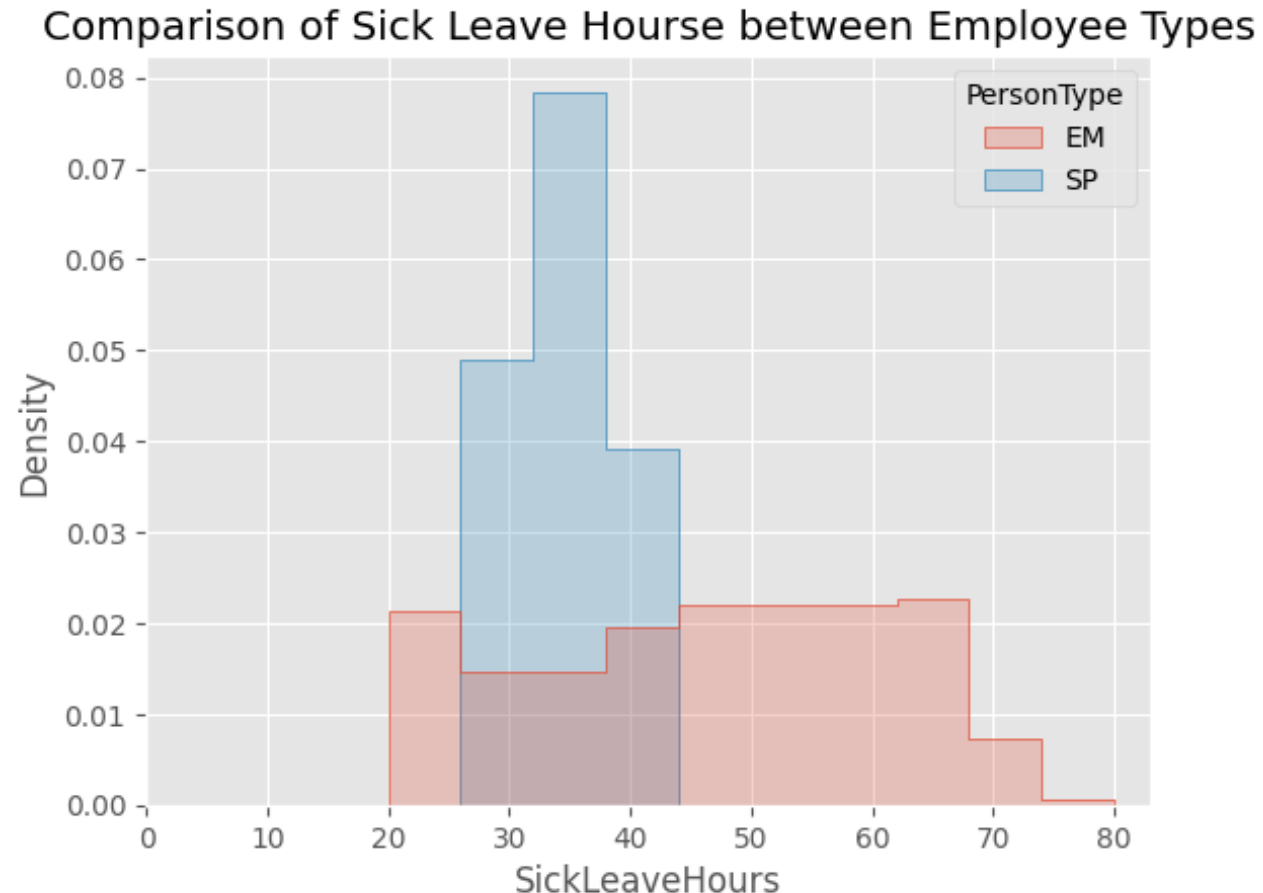
- Almost uniform distribution of sick leave hours
- Tapers off after 70 hours
- Everyone has taken at least 20 sick leave hours

# Distribution of Sick Leave Hours for EM and SP Employees



- Both distributions relatively symmetric and no outliers
- Noticeable difference in spread of sick leave hours between EM and SP
- Suggest SP employees take fewer sick leave hours!
- Is there a difference in company culture around taking sick leave?

# Normalised Histograms



- EM employees has nearly Uniform distribution
- There is a 93.04% chance of EM employee taking 20 - 70 hours of sick leave
- 94.12% chance of SP employing of taking 30 - 40 hours of sick leave

5. What is the relationship  
between store trading  
duration and revenue?

---

# How to Compare Stores

## Store Trading Duration:

The length of time a shop has been actively conducting sales



- First attempt at visualising
- We see there are common trading durations so...

# Group By Trading Duration

To compare

stores:

```
df["SalesPerDay"] = df["TotalSales"] / df["TradingDurationInDays"]
```

Group by trading duration to get Average SalesPerDay:

	TradingDurationInDays	SalesPerDay
1	89	21.583135
2	90	12.369644
3	91	13.464329
4	92	27.544248
5	181	28.719212



# Hypothesis Test

---

## **Null Hypothesis:**

There is no relationship between Average Trading Duration and Sales Per Day

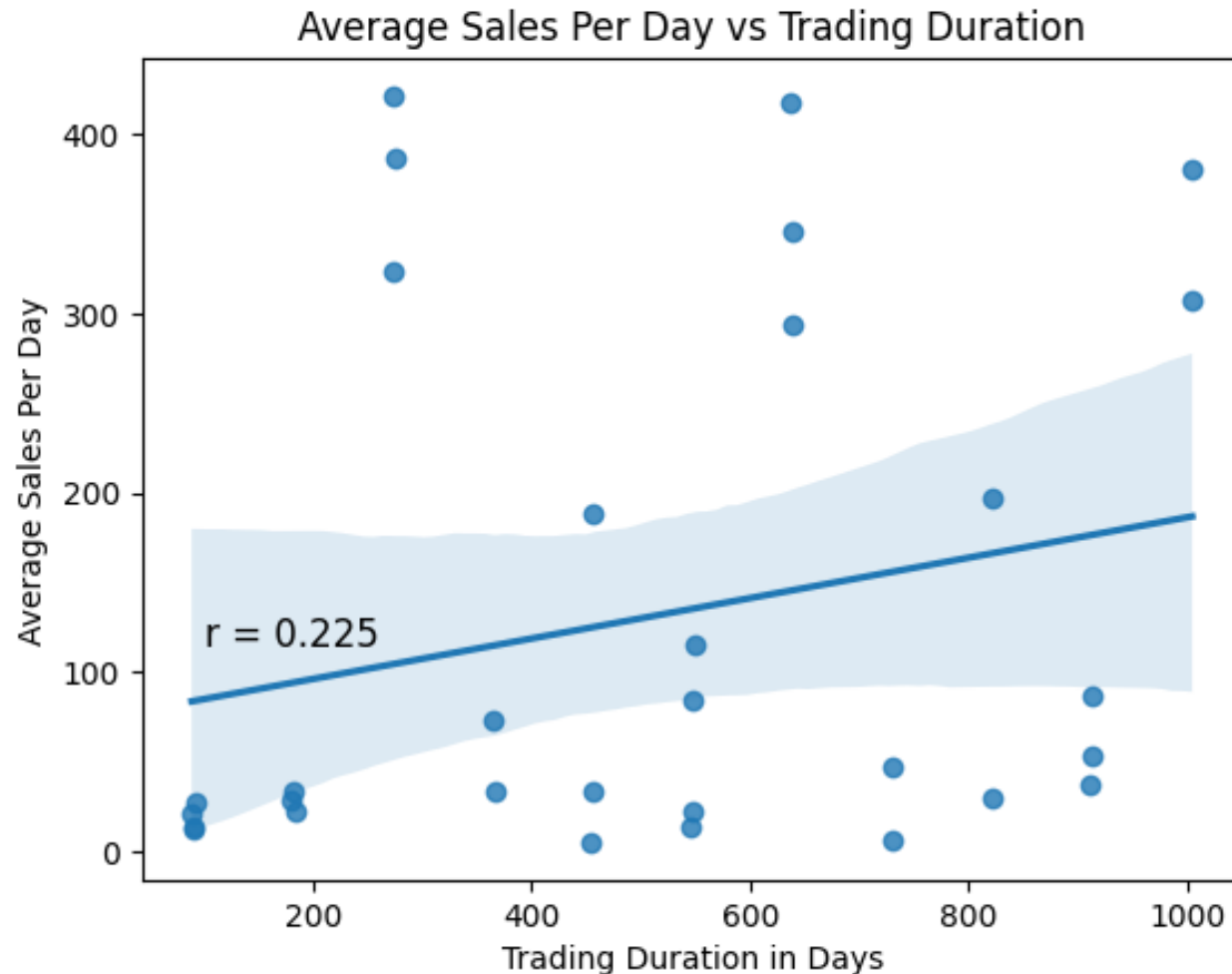
## **Alternative Hypothesis:**

There is a relationship between Average Trading Duration and Sales Per Day

## **What's our intuition?**

As Trading Duration increase, shops are more established which could lead to an increase in Sales Per Day

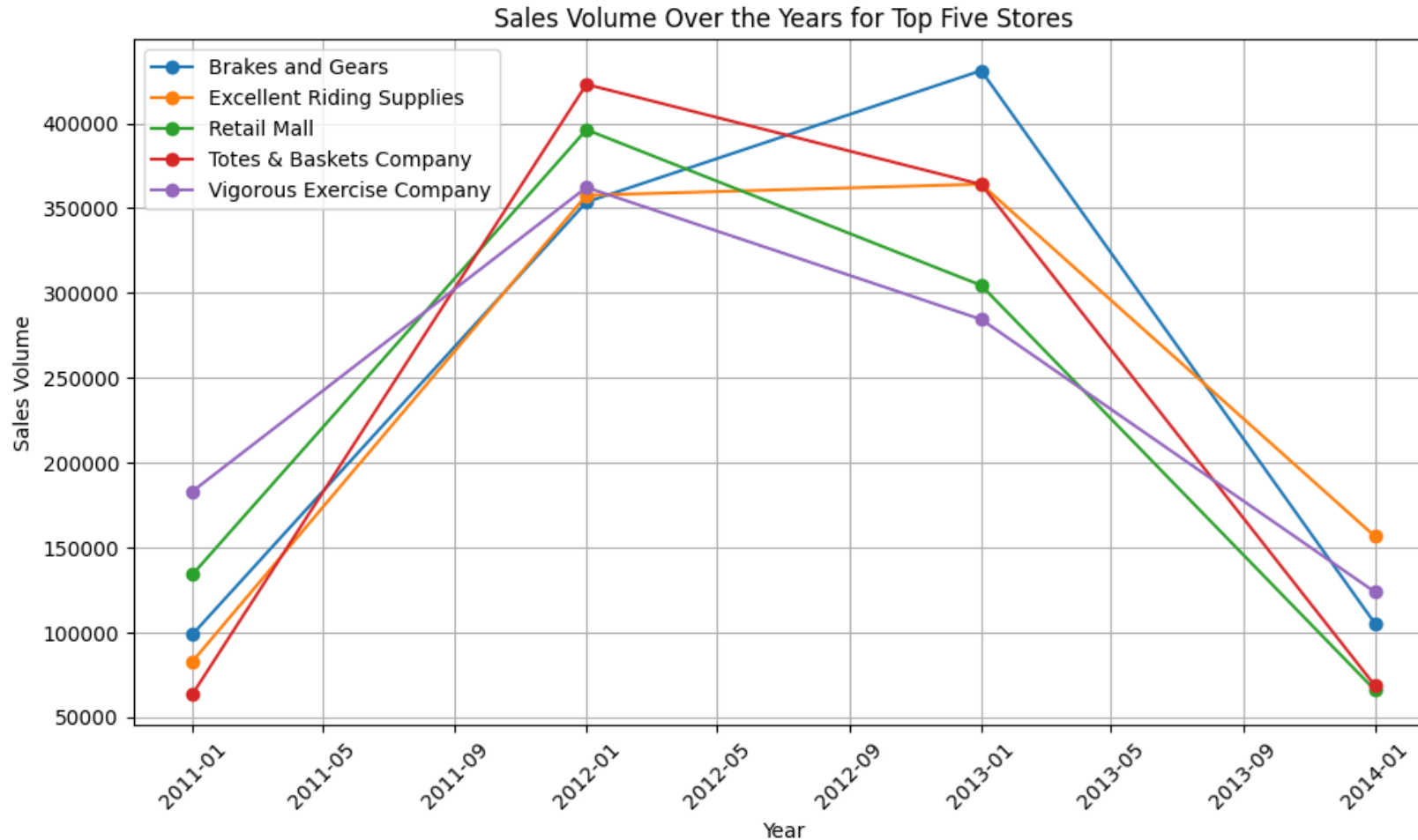
# Average Sales Per Day Compared to Trading Duration



- Sometimes it's useful to draw a best fit line, not this time
- Correlation coefficient shows very little relationship
- Cannot reject the Null Hypothesis at 5% significance level

```
PearsonRResult(statistic=0.22544793681527714, pvalue=0.22268021032551247)
```

# Volume of Sales Over Years

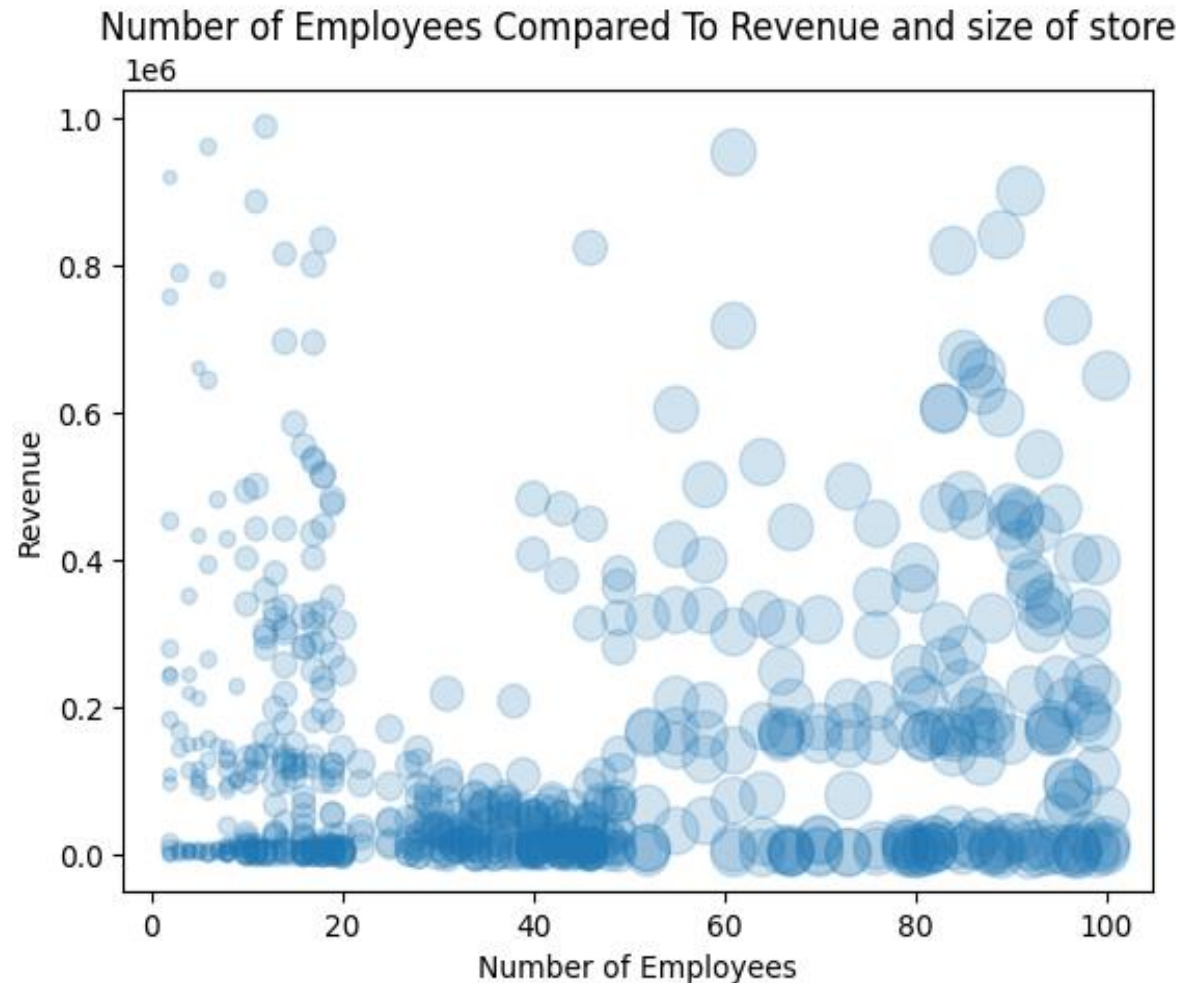


Q6) What is the relationship between the size of the stores, number of employees and revenue?

---

# First Attempt at Visualisation

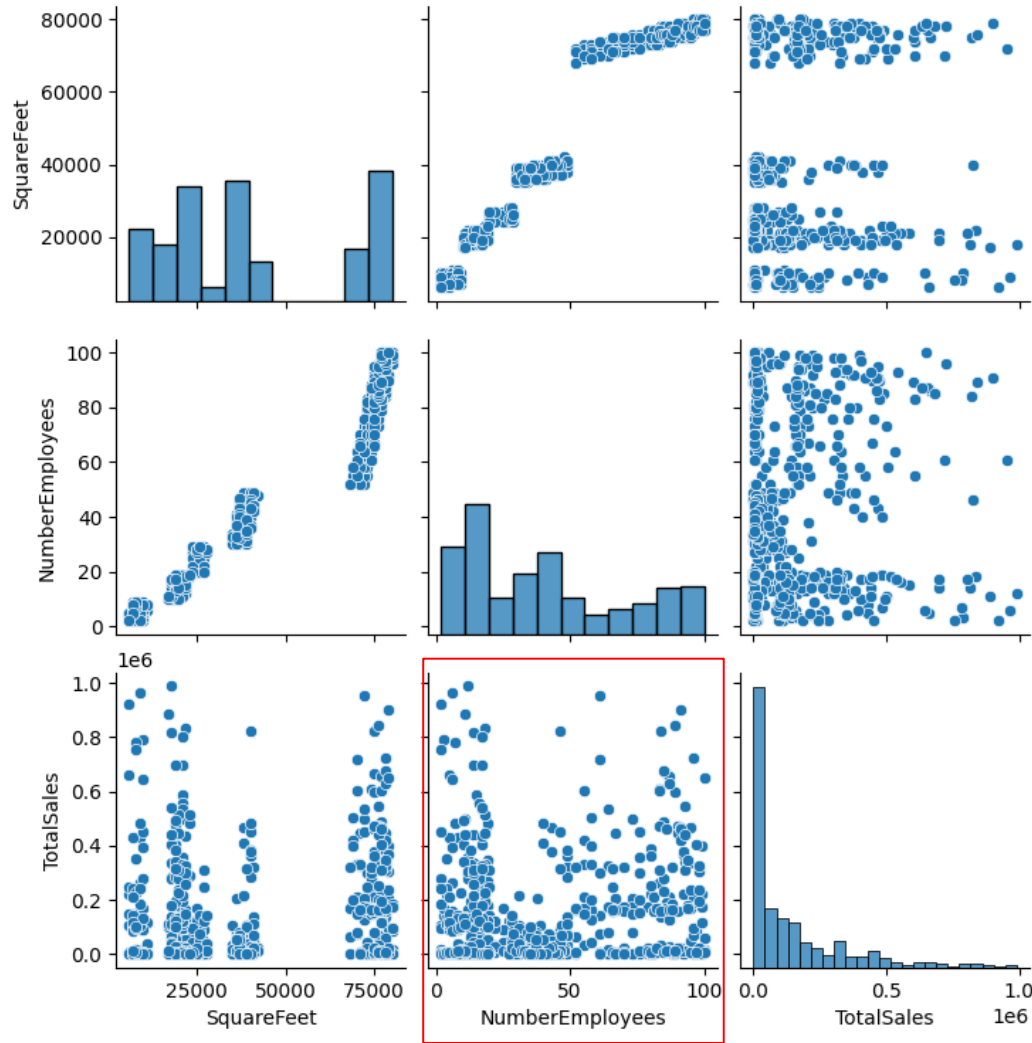
---



We have three continuous variables, so there are two options:

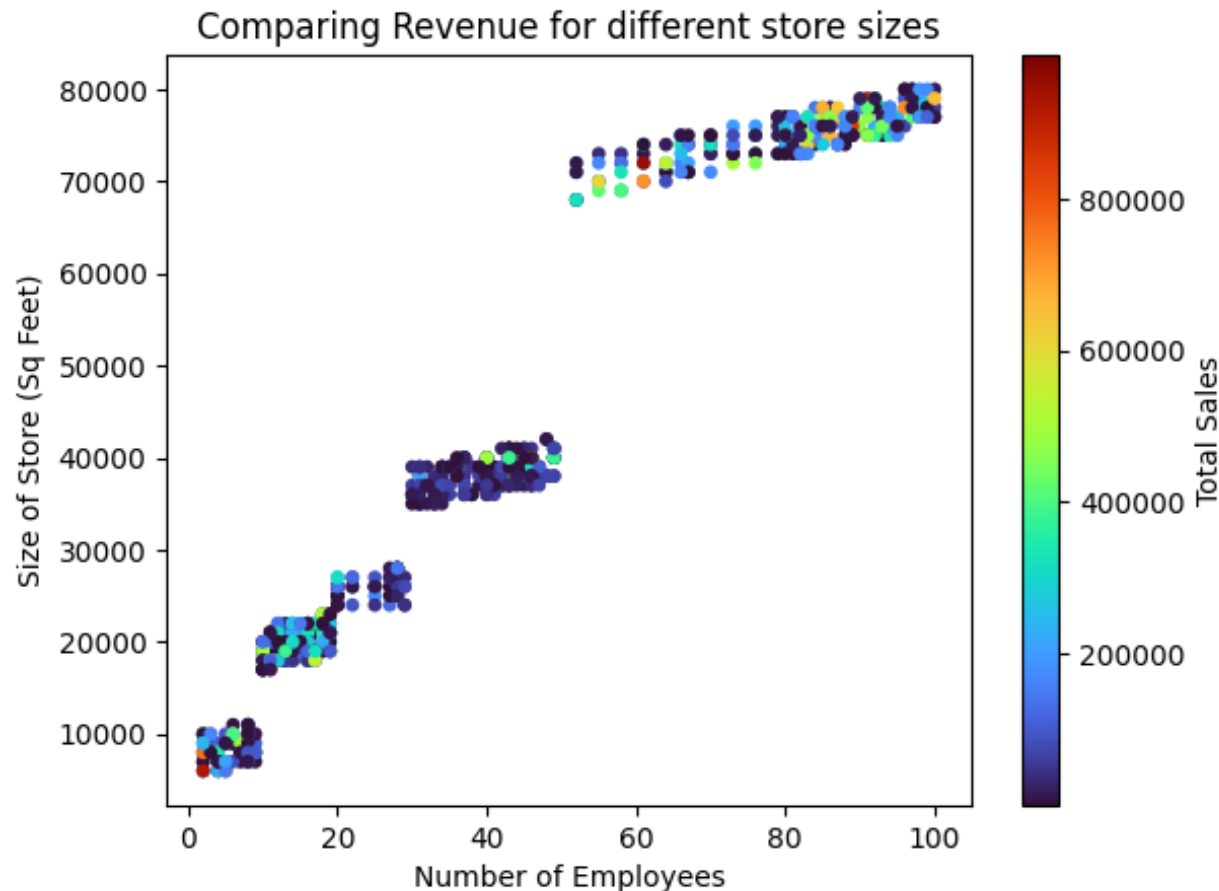
1. Use marker size to indicate store size
1. Use colour

# Is There Any Meaningful Relationship?



- No clear patterns between number of employees and total sales
- Bottom right: Most Total sales below \$42,000

# Sales Amounts Using Colour



- The number of employees for similar sized stores are clustered
- Similar sized stores require similar amount of employees
- No clear trend in Total Sales

# Hypothesis Test

---

## **Null Hypothesis:**

There is no relationship between number of employees and size of store (sqft)

## **Alternative Hypothesis:**

There is a positive relationship between number of employees and size of store (sqft)

## **What's our reasoning?**

Larger stores require more employees. These variables should be linked.



# Correlation Heat Map



- Focus on TotalSales column:
- Little to no evidence of a relationship between size of store or number of employees and TotalSales
- AdventureWorks should look for other relationships to affect sales

# Thank you