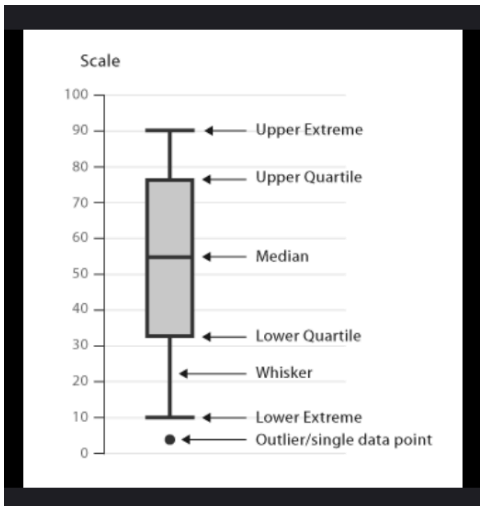
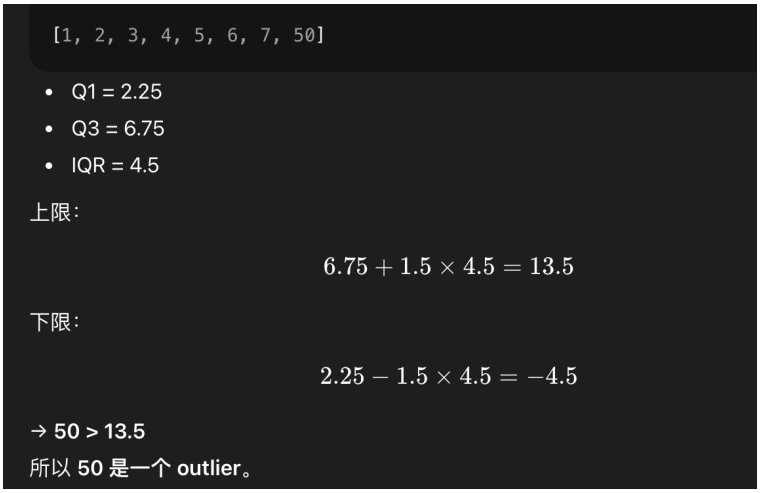
recap

Linear Regression sheet

- **Box Plot**
  A box plot identifies outliers as any values falling below Q1−1.5 × IQR or above Q3+1.5 × QR, where IQR (Interquartile Range) is the difference between Q3 and Q1.





1. Linearity

Plain English:

“The relationship between X and Y should look like a straight line, not curved.”

2. Independence

Plain English:

“Each observation should not depend on the others.”

3. Homoscedasticity (constant variance)

Plain English:

“The spread of residuals should be roughly the same everywhere — no funnel shape.”

4. Normality of residuals

Plain English:

“The errors should roughly follow a bell-shaped curve.”

5. No multicollinearity（only for multiple regression）

Plain English:

“The predictors should not be duplicates of each other.”

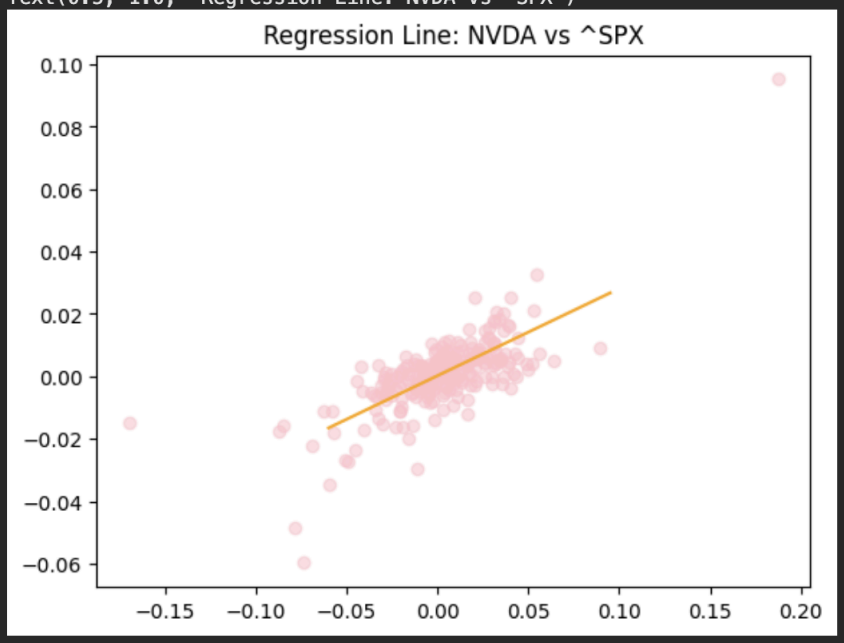
QQ Plot

A QQ plot compares the quantiles of my residuals with those of a normal distribution.
If the dots follow a straight line, I’m happy with the normality assumption.

Systematic curvature or S-shape means non-normal residuals, and a few points far away from the line are potential outliers or influential observations.

report trials

```
Text(0.5, 1.0, 'Regression Line: NVDA vs ^SPX')
```



[29]
✓ 0s

```
from sklearn.metrics import r2_score
r2 = r2_score(y, model.predict(x))
print("R² =", r2)
```

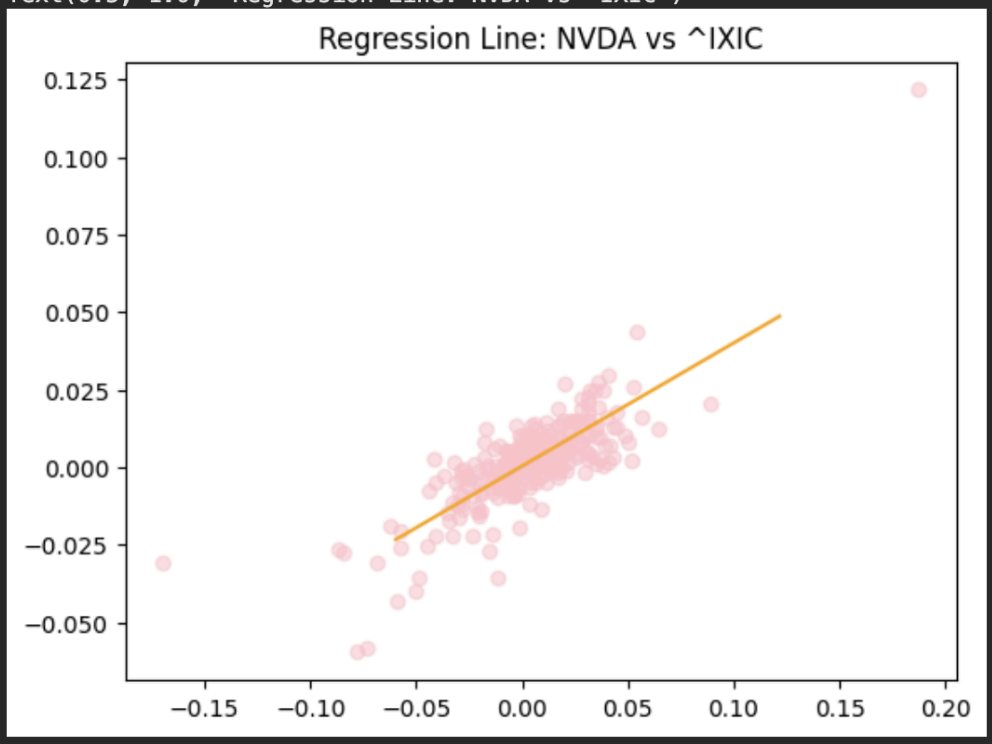
```
R² = 0.5332215598820924
```

Nvda vs SPX

https://colab.research.google.com/drive/1Mha-zHA-I4zjNJ3evLmqUznP9TFGws3f#scrollTo=JCTpiyPrrX75

```
plt.title(f'Regression Line: {company} vs {market}')
```

```
Text(0.5, 1.0, 'Regression Line: NVDA vs ^IXIC')
```



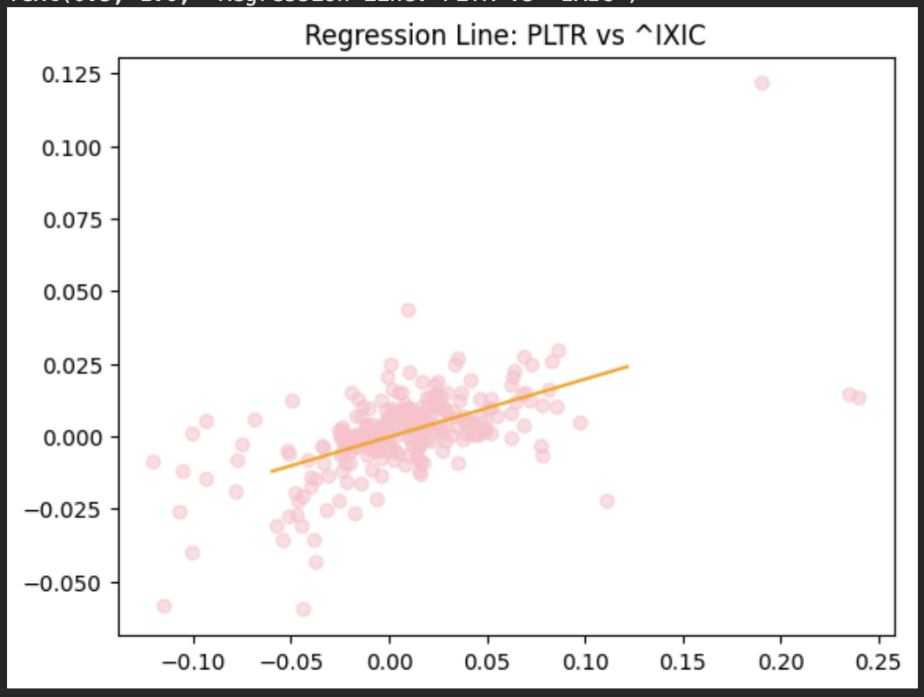

```
from sklearn.metrics import r2_score
r2 = r2_score(y, model.predict(x))
print("R² =", r2)
```

```
R² = 0.6460035366671779
```

Nvda vs IXIC

```
plt.title(f'Regression Line: {company} vs {market}')
```

```
Text(0.5, 1.0, 'Regression Line: PLTR vs ^IXIC')
```




```
from sklearn.metrics import r2_score
r2 = r2_score(y, model.predict(x))
print("R² =", r2)
```

```
R² = 0.3399769626695659
```

PLTR vs IXIC

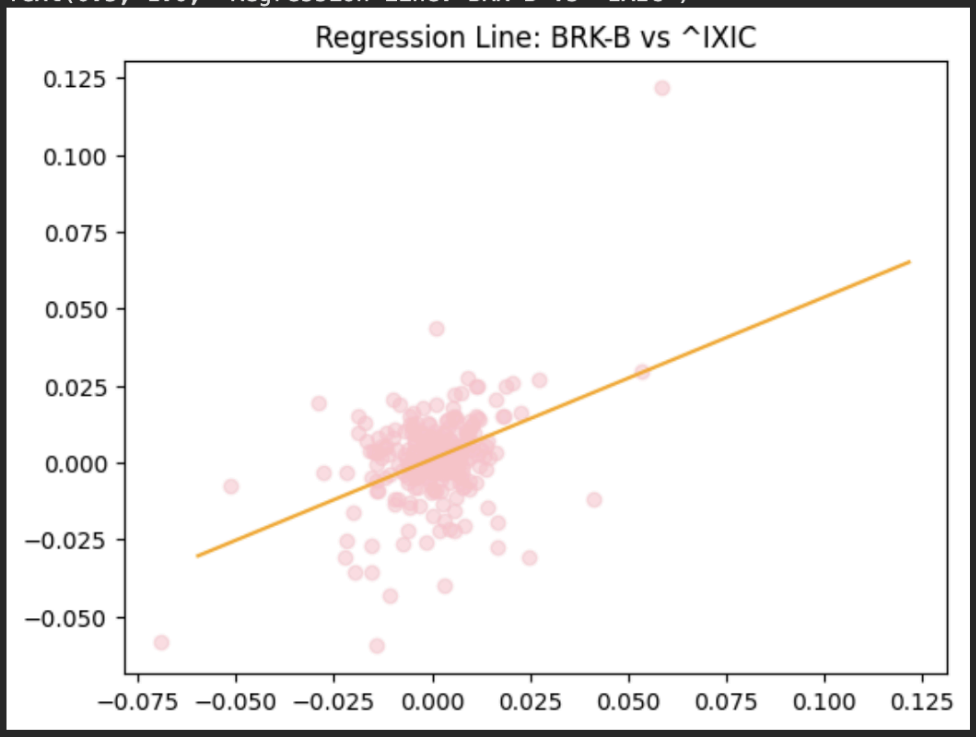
```
Text(0.5, 1.0, 'Regression Line: TSLA vs ^IXIC')
```


```
from sklearn.metrics import r2_score
r2 = r2_score(y, model.predict(x))
print("R² =", r2)
```

```
R² = 0.4984722554092955
```

Tesla vs IXIC

```
Text(0.5, 1.0, 'Regression Line: BRK-B vs ^IXIC')
```

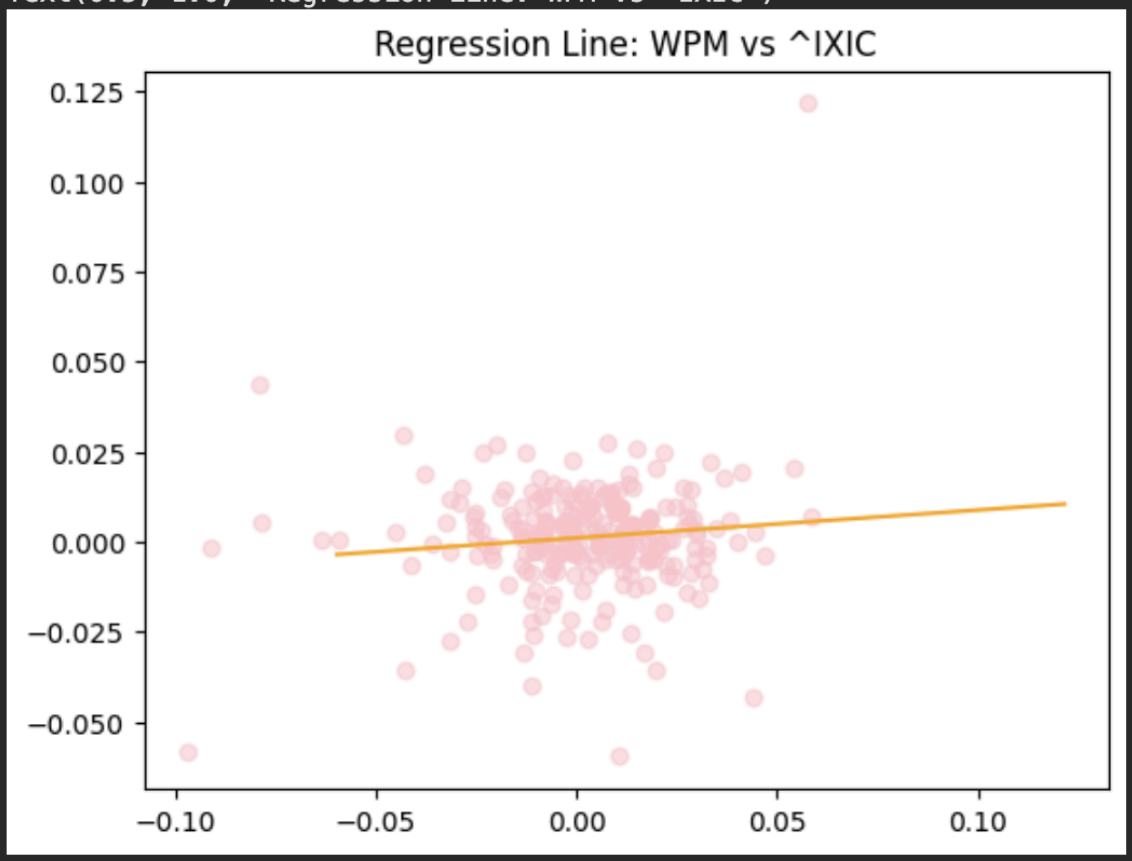



```
from sklearn.metrics import r2_score
r2 = r2_score(y, model.predict(x))
print("R² =", r2)
```

```
R² = 0.17312166554741015
```

BRK-B vs IXIC

```
Text(0.5, 1.0, 'Regression Line: WPM vs ^IXIC')
```




```
from sklearn.metrics import r2_score
r2 = r2_score(y, model.predict(x))
print("R² =", r2)
```

```
R² = 0.012706937751408742
```

WPM vs IXIC

report template

📄 Linear Regression Analysis Report — Template

---

1. Introduction

In this report, I use simple linear regression to study the relationship between major market indices and selected individual stocks.
The goal is to:

- Understand how strongly each stock moves with the overall market
- Compare high-beta, high-R^2 tech names (NVDA, TSLA, PLTR) with low-R^2 stocks (e.g., WPM)
- Interpret α (alpha) and β (beta)
- Evaluate model fit and outliers
- Provide intuition on why different stocks show different market sensitivities

The analysis uses daily price data from **startdatestart datestartdate to enddateend dateenddate**.

---

2. Data & Methodology

2.1 Data Sources

- Market indices: **S&P 500 (GSPC)** and **NASDAQ Composite (IXIC)**
- Stocks: **NVDA, TSLA, PLTR (high market sensitivity)**
- Comparison stock: **WPM (low market sensitivity)**

- Data frequency: **Daily close-to-close returns**

2.2 Return Calculation

rt=Pt−Pt−1Pt−1r_t = \frac{P_t - P_{t-1}}{P_{t-1}}rt=Pt−1Pt−Pt−1

2.3 Linear Regression Model

For each stock, we estimate:

rstock,t=α+βrmarket,t+εtr_{\text{stock}, t} = \alpha + \beta r_{\text{market}, t} + \varepsilon_trstock,t=α+βrmarket,t+εt

Where:

- **β** = sensitivity to market (systematic risk)
- **α** = abnormal return after accounting for market movement
- **R^2** = how much of the stock's movement the market explains

Market proxies used:

- **IXIC** (NASDAQ)
- **GSPC** (S&P500)

---

3. Results

3.1 Regression Results for Tech Stocks (High R^2)

3.1.1 NVDA vs IXIC

- **β:** fillfillfill
- **α:** fillfillfill

- **R^2:** fillfillfill

Interpretation:

- High β indicates NVDA moves strongly with the NASDAQ.
- High R^2 (e.g., >0.70) indicates the NASDAQ explains most of NVDA's daily price movement.
- This makes sense because NVDA is a major component of both Nasdaq and S&P500 and is tech-driven.

3.1.2 TSLA vs IXIC

- **β:** fillfillfill
- **α:** fillfillfill
- **R^2:** fillfillfill

Interpretation:

- TSLA also shows high sensitivity to tech sentiment and macro growth expectations.
- R^2 generally lower than NVDA but still high.

3.1.3 PLTR vs IXIC

- **β:** fillfillfill
- **α:** fillfillfill
- **R^2:** fillfillfill

Interpretation:

- PLTR's business is more cyclical and growth-driven, so its performance aligns strongly with Nasdaq movement.

3.2 Low-R^2 Example: WPM vs IXIC

- **β:** fillfillfill
- **α:** fillfillfill
- **R^2:** fillfillfill

Interpretation:

- WPM is a **precious metals royalty company**, so it behaves more like gold/silver rather than tech or macro growth.
- Therefore the NASDAQ explains almost none of its movement → **low R^2**.
- This serves as a good contrast to tech stocks.

---

4. Visualizations

Include:

4.1 Time Series Plots

- Market vs stock returns
- Overlay NVDA / TSLA / PLTR with IXIC

4.2 Scatter Plots with Regression Lines

- Stock returns vs index returns
- Visually shows slope (β) and tightness of fit (R^2)

4.3 Boxplots

- To detect outliers in stock returns

4.4 Residual Diagnostics

- Residual vs fitted plot
- Q–Q plot
- R^2 comparison bar chart

---

5. Interpretation & Discussion

5.1 Why Tech Stocks Have High R^2

- NASDAQ is tech-heavy
- NVDA/TSLA/PLTR have strong macro-growth beta
- These companies react to interest rates, AI sentiment, and risk-on behavior
- Makes their daily returns highly correlated with IXIC

5.2 Why WPM Has Low R^2

- Metal royalty companies follow commodity cycles
- Gold/silver driven by inflation, Fed policy, and geopolitical uncertainty
- Not related to tech market movements
- Therefore low β and low R^2 vs IXIC

5.3 Alpha Interpretation

- Positive $\alpha \rightarrow$ stock tends to outperform after adjusting for market movement

- Negative α → underperformance
- Most daily α values for large stocks are small, but meaningful over time

---

6. Conclusion

- Regression confirms tech-heavy stocks (NVDA, TSLA, PLTR) exhibit strong correlation with Nasdaq (IXIC), shown by high β and high R^2.
- Non-tech commodity-related stocks like WPM show low sensitivity to NASDAQ and low R^2.
- Linear regression effectively reveals how much a stock's movement is driven by the overall market vs. sector-specific factors.

---

7. Appendix

Include:

- Full regression tables
- Code snippet
- Diagnostics (QQ plot, residual plot, Cook's Distance, etc.)