GROUP 2

1. Name

Shiny Alexander Khootozhathil

Rishav Mondal

Dhruv Pancholi

2. Interested Industry's 4 digits SIC Code and Name

National Commercial Banks, SIC Code: 6021

3. Description of the Interested Industry

The National Commercial Banks industry in the USA, represented by SIC Code 6021, encompasses institutions offering varied banking services on a national scale, accepting deposits, providing loans, and facilitating financial transactions. National commercial banks play a significant role in the economy by serving as intermediaries for capital allocation and the provision of financial services. This sector comprises 81,450 businesses employing about 2,145,076 individuals. [1]

4. Company in the Interested Industry and its SEC EDGAR Page

Shiny- JPMORGAN CHASE & CO

SEC EDGAR Page :JPMORGAN CHASE & CO - SEC.gov

Rishav CAPITAL BANK FINANCIAL CORP

SEC EDGAR Page : EDGAR Search Results (sec.gov)

Dhruv- Bank of America-

SEC EDGAR Page https://www.sec.gov/edgar/browse/?CIK=0000070858&owner=exclude

5. Description of the Selected Company

- **Shiny: JPMORGAN CHASE & CO**

  JPMorgan Chase & Co. is a global financial powerhouse headquartered in New York City, USA. Its CEO, Jamie Dimon, oversees a diverse range of financial services, including retail and investment banking, asset management, and commercial banking. JPMorgan

---

[1] https://www.ibisworld.com/classifications/us-sic/6021/national-commercial-banks/

Chase, which has a rich history dating back to the nineteenth century, is known for its innovation and global reach, serving millions of customers worldwide. As one of the world's largest and most profitable banks, it plays a critical role in the international financial landscape.[2]

- **Rishav: CAPITAL BANK FINANCIAL CORP**

As of the first quarter of 2017, Capital Bank Financial Corporation has 193 branches and $10 billion in assets. Its headquarters are in Charlotte, North Carolina. Former vice chairman of the Bank of America Gene Taylor served as CEO; Chris Marshall, a former executive of the Bank of America and CFO of Fifth Third Bank, served as CFO; and R. Bruce Singletary, a former executive of the Bank of America, served as Chief Risk Officer. Kenneth Posner, a former speciality finance research analyst at Morgan Stanley, oversaw investor relations and corporate strategy. After a successful initial public offering (IPO) in September 2012, Capital Bank was traded on the NASDAQ under the symbol CBF until First Horizon National Corporation purchased the company in December 2017.

- **Dhruv: Bank of America**

The Bank of America was founded by Amadeo Pietro Giannini in 1904 based in San Francisco, United States. This financial firm has developed over the course of its existence into the second- largest banking organization in the United States, in charge of a sizeable chunk of deposits in American banks totaling about 10.73%.[3]

6. Interested Economic Indicator and its FRED Page
Shiny: Recession Probability

FRED Page: https://fred.stlouisfed.org/series/RECPROUSM156N

Rishav: Existing Home Sales

FRED Page: https://fred.stlouisfed.org/series/EXHOSLUSM495S#0

Dhruv: Interest Rate

FRED Page: - https://fred.stlouisfed.org/series/DFF

7. Description of the Interested Economic Indicator
- Shiny's Description of Recession Probability

---

[2]https://en.wikipedia.org/wiki/JPMorgan_Chase#:~:text=JPMorgan%20Chase%20%26%20Co,market%20capitalization%20as%20of%202023

[3] https://en.wikipedia.org/wiki/Bank_of_America

The recession probability metric evaluates the chances of a recession using diverse methods. It incorporates financial indicators like changes in the yield curve and the difference between various interest rates. It also includes key economic indicators, such as consumer confidence and critical macroeconomic factors like inflation and joblessness rates. To predict the likelihood of a notable rise in unemployment in the near and distant future, models, often logistic regression, are employed. This comprehensive approach, which merges financial, economic, and market information, highlights the complexities involved in forecasting economic downturns.[4]

- Rishav's Description of Existing Home Sales

An essential tool for assessing the state of the housing market and general economic activity is can spot possible changes in the housing sector that could have wider effects on the whole economy. The strength of consumer borrowing and spending patterns is reflected in the Existing Home Sales report, which is regarded as a leading economic indicator. Interest rates can be influenced by past home sales, which are a good way to measure consumer spending. Through the examination of the Existing Home Sales report, investors and economists can spot possible changes

- Dhruv's Description of Interest Rate

Federal funds interest rate refers to the target rate that is generally set by the Federal Open Market Committee every eight years. Commercial banks usually follow this rate in order to borrow or lend money to other commercial banks. [5]

8. Usage of the Interested Economic Indicator
- Shiny's Usage of Recession Probability

The recession probability indicator is primarily used to predict the likelihood of a recession, serving as a crucial tool for economists and financial market analysts. It's employed when assessing economic conditions, especially during periods of uncertainty or market instability. This indicator helps in forecasting economic downturns by analyzing a combination of financial variables, leading economic indicators, and macroeconomic data. Its use is vital in guiding economic policy, investment decisions, and financial planning, aiming to preempt and mitigate the impacts of potential economic recessions.[6]

- Rishav's Usage of Existing Home Sales

The monthly sales of previously owned single-family homes, condos, and cooperatives are measured using existing home sales. Every month, a statistical report is released by the National

---

[4]https://www.federalreserve.gov/econres/notes/feds-notes/financial-and-macroeconomic-indicators-of
- recession-risk-20220621.html

[5] https://www.investopedia.com/terms/f/federalfundsrate.asp

[6] https://www.newyorkfed.org/research/capital_markets/ycfaq#/

Association of Realtors (NAR).Current home sales have an impact on borrowing rates and are regarded as a gauge of consumer spending. A robust increase in current house sales is favorable for the dollar. It is dollar bearish if existing house sales fall over a few months.Data on previous house sales can also be used to identify trends in real estate markets. For instance, if prospective homebuyers are being discouraged by increased mortgage interest rates and property values. It can offer a quick overview of the state of the industry, including the trends and problems that the industry is facing on a national and local level.[7]

- Dhruv's Usage of Interest Rate

Federal Funds Rate is an interest rate in which bank trades money with each other overnight. When a depository institution has excess balances in its reserve account it lends money to another bank. The rate at which borrowing institutions pay to the lending institutions is determined by both the parties, and the overall average rates of all these agreements is known as federal funds rate.[8]

9. Research Question or Hypothesis

Shiny's Research Question

How does the recession probability indicator affect JPMorgan Chase's net income?

Evidence

Ni and Recession Probability
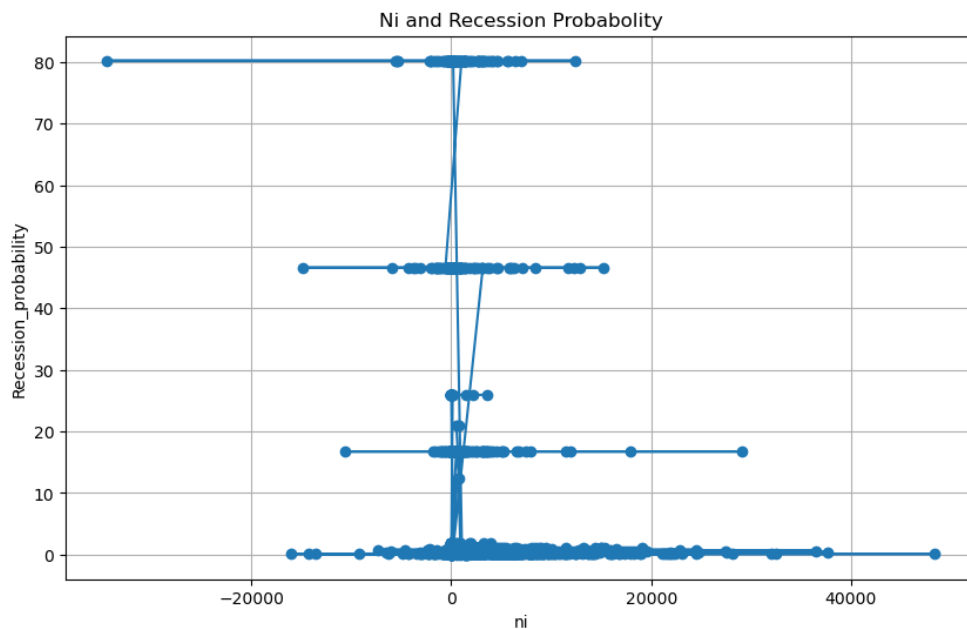
Summary:
It is a well-known fact that a rise in the recession probability indicator is typically linked to a

---

[7] ibisworld.com, advisorperspectives.com, nar.realtor, babypips.com
[8] https://www.investopedia.com/terms/f/federalfundsrate.asp

possible decline in economic activity, which can have an effect on banks' financial health. Because of things like higher loan loss provisions, less demand for loans, and lower interest income, this can result in a decline in net income.[9]

Expected Relationship

As you can see, the recession probability has a negative correlation with NI variables. This means that as the recession probability increases, the values of this variable tend to decrease. This is consistent with the idea that a recession is a period of economic decline that is characterized by lower levels of economic activity.

Rishav's Research Question

How is housing sales getting affected when there's a rise in recession or federal funds interest rate?

Ans.

We were able to provide and experimental evaluation for this research question because of lack of data in the **'Housing Units Sold'** FRED dataset, which is causing underfitting and producing skewed results. Hence I have to provide theoretical answers.

Impact of Federal Funds Rate:

Lower Federal Funds rates generally have the ability to boost economic activity by lowering the cost of borrowing. This could stimulate the home market by raising the demand for mortgages.

While, conversely if Federal Funds rates are high, then it may lead to decrease economic activity by increasing the cost of borrowing. This could stimulate the home market by lowering the demand for mortgages.

Impact of Recession Probability:

A high likelihood of recession may induce economic instability, resulting in decreased consumer assurance and expenditure. The housing market may suffer as a result, since prospective buyers might put off purchases during shaky economic times.

Conversely, a low likelihood of recession may induce economic growth and stability, resulting in increased consumer assurance and expenditure. The housing market may become better as a result, since prospective buyers might purchases during stable economic times.

Combined Effect of the variables:

The house market may suffer if there is a significant risk of a recession and a high Federal Funds rate, which indicates higher borrowing costs. In an unstable economy, homebuyers might have to deal with greater costs and be wary of making large financial commitments.

On the other hand, the housing market may benefit if both the Federal Funds rate, which indicates lower borrowing costs, and the likelihood of a recession are low. During an economic moment of stability and confidence, homebuyers may reap the benefits of reduced expenses and exhibit a greater willingness to make substantial financial commitments.

---

[9] https://www.bloomberg.com/news/articles/2023-01-21/jpmorgan-model-shows-recession-odds-fall-sharply-across-markets

Dhruv's Research Question

Is there a direct impact on profitability of banks when there's a fluctuation in federal funds rate?

```
                          OLS Regression Results
================================================================================
Dep. Variable:                   revt   R-squared:                       0.874
Model:                            OLS   Adj. R-squared:                  0.873
Method:                 Least Squares   F-statistic:                     1665.
Date:                Tue, 05 Dec 2023   Prob (F-statistic):           3.41e-217
Time:                        09:58:22   Log-Likelihood:                 -4891.6
No. Observations:                 485   AIC:                             9789.
Df Residuals:                     482   BIC:                             9802.
Df Model:                           2
Covariance Type:            nonrobust
================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
Intercept              8.6497      1.125      7.689      0.000       6.439      10.860
ni                     4.8273      0.087     55.210      0.000       4.655       4.999
Housing_Units_Sold    -0.0012      0.000     -6.395      0.000      -0.002      -0.001
Federal_Funds       3118.6920    405.589      7.689      0.000    2321.751    3915.633
Recession_probability -17.6752      2.299     -7.689      0.000     -22.192     -13.159
================================================================================
Omnibus:                      396.309   Durbin-Watson:                   1.866
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            12465.522
Skew:                           3.276   Prob(JB):                         0.00
Kurtosis:                      26.957   Cond. No.                     2.88e+23
================================================================================
```

- **R-Squared:**

1. The R-squared is 0.874, which means that 87.4% of dependent variable (revt) can be explained by the independent variables (ni,housing units sold, Federal_Funds, and Recession_probability).
2. A higher R-squared means a better fit of the model to the data.

- **P-Value(F-Statistic):** In this case, the F-statistic is 1665. with a p-value of 3.41e-217, which indicates that the overall model is statistically significant.

- **P-Value(economic variables):** In this case, the t-statistic is 7.689 for **'federal funds'** with a p-value of 0.000, which indicates that the relationship of the independent variables with the dependent variable is statistically significant .

10. Data

a. Brief description of Compustat and FRED:

Compustat

Compustat is a financial and market database for thousands worldwide companies, covering data from 1950. It's a go-to source of information for experts in finance, shareholders, and researchers. [10]

Compustat is a global database of fundamental financial and market information on both operating and inactive global companies. Compustat is a leading computerised source of market intelligence and company data—and one of the oldest: Standard & Poor's (S&P) has been producing it since 1962.[11]

FRED:

The Federal Reserve Bank of St. Louis created and maintains FRED (**Federal Reserve Economic Data**), a free online database of economic data from various sources. It gives users tools for interacting with, displaying, and disseminating data. [12]

FRED is a database of US economic time series that allows users to access over 800,000 data sets that measure specific phenomena over time.[13]

11. Data Preparation(Code)

a. Description of how data is imported

---

[10] https://www.refinitiv.com/en/financial-data/company-data/fundamentals-data/standardized-fundamentals/sp-compustat-database

[11] https://www.investopedia.com/terms/c/capital-iq.asp

[12] https://fredhelp.stlouisfed.org/fred/about/about-fred/what-is-fred/

[13] https://smartasset.com/financial-advisor/federal-reserve-economic-data-fred

Visualising the data ¶

```python
Nat_Comm_Bank = conn.raw_sql('''select cik, gvkey, datadate, conm, revt, ni from comp.funda
                    where sich=6020 and datadate>='01/01/1990' and datadate<='12/31/2023'
                    and  datafmt = 'STD' and consol = 'C'and indfmt = 'INDL'
                    ''', date_cols=['datadate'])
Nat_Comm_Bank['year']=pd.DatetimeIndex(Nat_Comm_Bank['datadate']).year
Nat_Comm_Bank['month']=pd.DatetimeIndex(Nat_Comm_Bank['datadate']).month
```

```python
Nat_Comm_Bank
```

	cik	gvkey	datadate	conm	revt	ni	year	month
0	0000763901	002002	2009-12-31	POPULAR INC	2751.498	-573.919	2009	12
1	0000763901	002002	2010-12-31	POPULAR INC	2595.637	137.401	2010	12
2	0000763901	002002	2011-12-31	POPULAR INC	2497.778	151.325	2011	12
3	0000763901	002002	2012-12-31	POPULAR INC	2218.047	245.275	2012	12
4	0000763901	002002	2013-12-31	POPULAR INC	2128.709	599.327	2013	12
...	...	...	...	...	...	...	...	...
10219	0000931061	268025	2018-12-31	ABBEY NATL TREASURY SVCS PLC	21.029	24.470	2018	12
10220	0000931061	268025	2019-12-31	ABBEY NATL TREASURY SVCS PLC	88.337	-8.873	2019	12
10221	0001471055	285313	2007-12-31	BANCO SANTANDER BRASIL -ADR	9548.899	1069.702	2007	12
10222	0001471055	285313	2008-12-31	BANCO SANTANDER BRASIL -ADR	10876.378	1026.055	2008	12
10223	0001471055	285313	2009-12-31	BANCO SANTANDER BRASIL -ADR	24871.241	3160.749	2009	12

10224 rows × 8 columns

FRED DATA

```python
mykey='f5d244832c04e7d91938cc305d3a429d'
```

```python
fred = Fred(api_key=mykey)
```

```python
Exstng_Home_Sls = 'EXHOSLUSM495S'#Existing home sales
Reces_Prob='RECPROUSM156N' #Recession PRobability
Fed_Funds = 'DFF' #Federal Funds Effective Rate
```

```python
Hme_sls = fred.get_series(Exstng_Home_Sls, obsevation_start = '1990-01-01')
FF_IR=fred.get_series(Fed_Funds,observation_start='1990-01-01')
R_Prob=fred.get_series(Reces_Prob, observation_start='1990-01-01')
```

Using economic indicator keys to access FRED and the SIC code 6021 to access WRDS are the first steps in importing data from both sources. Utilizing these distinct identifiers, this procedure entails logging into each platform and querying the relevant data sets.

b. Primary key of data from WRDS and FRED individually

	year	Federal_Funds	Recession_probability	cik	gvkey	datadate	conm	revt	ni	month	Housing_Units_Sold
0	2022	1.692192	0.416667	0000763901	002002	2022-12-31	POPULAR INC	3096.073	1102.641	12	4.196667e+06
1	2022	1.692192	0.416667	0000046195	002005	2022-12-31	BANK OF HAWAII CORP	754.907	225.804	12	4.196667e+06
2	2022	1.692192	0.416667	0000036146	004685	2022-12-31	TRUSTMARK CORP	746.977	71.887	12	4.196667e+06
3	2022	1.692192	0.416667	0000798941	004690	2022-12-31	FIRST CITIZENS BANCSH -CL A	5105.000	1098.000	12	4.196667e+06
4	2022	1.692192	0.416667	0000073124	007982	2022-12-31	NORTHERN TRUST CORP	7751.700	1336.000	12	4.196667e+06
...	...	...	...	...	...	...	...	...	...	...	...
478	2023	4.998627	0.394000	0001144967	144535	2023-03-31	HDFC BANK LTD	24921.884	5600.999	3	4.158000e+06
479	2023	4.998627	0.394000	0001103838	223148	2023-03-	ICICI BANK LTD	17719.993	4144.591	3	4.158000e+06

Together, the 'year' and 'gvkey' columns make up the primary key in the combined dataset. Accordingly, every set of 'gvkey' and 'year' values uniquely identifies a single dataset entry. 'gvkey' and 'year' values cannot be combined in any two entries. As a result, the primary keys of the dataset are now "gvkey" and "year."

c. **Number of observations (WRDS and FRED individually)**

d. **WRDS (Wharton Research Data Services):**

```
]: Nat_Comm_Bank.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10224 entries, 0 to 10223
Data columns (total 8 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   cik       9823 non-null   object
 1   gvkey     10224 non-null  object
 2   datadate  10224 non-null  datetime64[ns]
 3   conm      10224 non-null  object
 4   revt      10212 non-null  float64
 5   ni        10218 non-null  float64
 6   year      10224 non-null  int64
 7   month     10224 non-null  int64
dtypes: datetime64[ns](1), float64(2), int64(2), object(3)
memory usage: 639.1+ KB
```

Total Entries: 10224

FRED (Federal Reserve Economic Data):

Rishav

```
]: Hme_sls
```

```
]: 2022-09-01          NaN
   2022-10-01    4440000.0
   2022-11-01    4120000.0
   2022-12-01    4030000.0
   2023-01-01    4000000.0
   2023-02-01    4550000.0
   2023-03-01    4430000.0
   2023-04-01    4290000.0
   2023-05-01    4300000.0
   2023-06-01    4160000.0
   2023-07-01    4070000.0
   2023-08-01    4040000.0
   2023-09-01    3950000.0
   2023-10-01    3790000.0
   dtype: float64
```

Home Sales: 14 Entries

Dhruv

```
16]: FF_IR
```

```
16]: 1990-01-01    7.97
     1990-01-02    8.54
     1990-01-03    8.37
     1990-01-04    8.29
     1990-01-05    8.20
                    ...
     2023-11-27    5.33
     2023-11-28    5.33
     2023-11-29    5.33
     2023-11-30    5.33
     2023-12-01    5.33
     Length: 12388, dtype: float64
```

Federal Funds:12388 Entries

Shiny

```
17]: R_Prob
```

```
17]: 1990-01-01    0.26
     1990-02-01    0.10
     1990-03-01    0.36
     1990-04-01    1.16
     1990-05-01    1.60
                   ...
     2023-06-01    0.38
     2023-07-01    0.10
     2023-08-01    0.14
     2023-09-01    0.36
     2023-10-01    2.22
     Length: 406, dtype: float64
```

Recession Probability: 406 Entries

Combined Dataset (HomeSales_FederalFunds_Recession Probability):

Inner merging with the housing sales FRED dataset

```
[34]: #merging the econ_var_Nat_com_1 dataset obtained above with the Housing_unites_sold dataset using 'left'
      econ_var_Nat_com_merged = pd.merge(econ_var_Nat_com_1, Hme_sls_group, on='year', how='inner')
      econ_var_Nat_com_merged#dataset with the economic variable 'Housing Units Sold'
```

[34]:

	year	Federal Funds	Recession probability	cik	gvkey	datadate	conm	revt	ni	month	Housing Units Sold
0	2022	1.692192	0.416667	0000763901	002002	2022-12-31	POPULAR INC	3096.073	1102.641	12	4.196667e+06
1	2022	1.692192	0.416667	0000046195	002005	2022-12-31	BANK OF HAWAII CORP	754.907	225.804	12	4.196667e+06
2	2022	1.692192	0.416667	0000036146	004685	2022-12-31	TRUSTMARK CORP	746.977	71.887	12	4.196667e+06
3	2022	1.692192	0.416667	0000798941	004690	2022-12-31	FIRST CITIZENS BANCSH -CL A	5105.000	1098.000	12	4.196667e+06
4	2022	1.692192	0.416667	0000073124	007982	2022-12-31	NORTHERN TRUST CORP	7751.700	1336.000	12	4.196667e+06
...	...	...	...	...	...	...	...	...	...	...	...
478	2023	4.998627	0.394000	0001144967	144535	2023-03-31	HDFC BANK LTD	24921.884	5600.999	3	4.158000e+06
479	2023	4.998627	0.394000	0001103838	223148	2023-03-31	ICICI BANK LTD	17719.993	4144.591	3	4.158000e+06
480	2023	4.998627	0.394000	0000067088	252940	2023-03-31	MITSUBISHI UFJ FINANCIAL GRP	43244.958	4512.706	3	4.158000e+06
481	2023	4.998627	0.394000	0001538263	170419	2023-06-30	HOMETRUST BANCSHARES INC	218.596	44.604	6	4.158000e+06
482	2023	4.998627	0.394000	0001335730	248136	2023-03-31	MIZUHO FINANCIAL GROUP INC	37671.024	4178.857	3	4.158000e+06

483 rows × 11 columns

Total Entries: 483

e. Steps related to changing the data frame (adding variables, aggregating data, etc.)

Aggregating the FRED datasets

Groupinng the average data according to the year

```python
Hme_sls_group = Hme_sls.groupby('year')['Housing Units Sold'].mean()
Hme_sls_group=Hme_sls_group.to_frame().reset_index()

gb_group=R_Prob.groupby('year')['Recession probability'].mean()
gb_group=gb_group.reset_index()

FF_group=FF_IR.groupby('year')['Federal Funds'].mean()
FF_group=FF_group.reset_index()
```

Rishav

```python
Hme_sls_group
```

	year	Housing Units Sold
0	2022	4.196667e+06
1	2023	4.158000e+06

Shiny

gb_group

	year	Recession probability
0	1990	20.931667
1	1991	12.460000
2	1992	0.101667
3	1993	0.115000
4	1994	0.008333
5	1995	0.153333
6	1996	0.065000
7	1997	0.015000
8	1998	0.090000
9	1999	0.030000
10	2000	1.780000
11	2001	25.935000
12	2002	0.881667
13	2003	0.321667
14	2004	0.133333
15	2005	0.958333
16	2006	0.101667
17	2007	1.120000
18	2008	80.221667
15	2009	46.610000
20	2010	0.070000
21	2011	0.091667
22	2012	0.143333
23	2013	0.060000
24	2014	0.051667
25	2015	0.773333
26	2016	0.248333
27	2017	0.066667
28	2018	0.166667
29	2019	0.613333
30	2020	16.728333
31	2021	0.101667
32	2022	0.416667
33	2023	0.394000

Dhruv

0]: `FF_group`

0]:

	year	Federal Funds
0	1990	8.096521
1	1991	5.685014
2	1992	3.521066
3	1993	3.021342
4	1994	4.206329
5	1995	5.834301
6	1996	5.300464
7	1997	5.461507
8	1998	5.350932
9	1999	4.972356
10	2000	6.237432
11	2001	3.878356
12	2002	1.666795
13	2003	1.126493
14	2004	1.350328
15	2005	3.217068
16	2006	4.965425
17	2007	5.017315
18	2008	1.924180
19	2009	0.158603
20	2010	0.176247
21	2011	0.101507
22	2012	0.140574
23	2013	0.107342
24	2014	0.068493
25	2015	0.133726
26	2016	0.393197
27	2017	1.002959
28	2018	1.834932
29	2019	2.157260
30	2020	0.372240
31	2021	0.078685
32	2022	1.692192
33	2023	4.998627

We sorted the entries in the dataset according to year. In the process, we determined the mean for every group, which allowed us to aggregate the data and produce average values for every year. This stage makes it possible to analyze the yearly trends in the data in a more efficient and perceptive manner.

12. Merging Data

a. Relation between datasets and the reason for selection (e.g., One-to-one, one-to-many, etc.)

```
: Nat_Comm_Bank = conn.raw_sql('''select cik, gvkey, datadate, conm, revt, ni from comp.funda
                                where sich=6020 and datadate>='01/01/1990' and datadate<='12/31/2023'
                                and  datafmt = 'STD' and consol = 'C'and indfmt = 'INDL'
                                ''', date_cols=['datadate'])
  Nat_Comm_Bank['year']=pd.DatetimeIndex(Nat_Comm_Bank['datadate']).year
  Nat_Comm_Bank['month']=pd.DatetimeIndex(Nat_Comm_Bank['datadate']).month
```

: Nat_Comm_Bank

	cik	gvkey	datadate	conm	revt	ni	year	month
0	0000763901	002002	2009-12-31	POPULAR INC	2751.498	-573.919	2009	12
1	0000763901	002002	2010-12-31	POPULAR INC	2595.637	137.401	2010	12
2	0000763901	002002	2011-12-31	POPULAR INC	2497.778	151.325	2011	12
3	0000763901	002002	2012-12-31	POPULAR INC	2218.047	245.275	2012	12
4	0000763901	002002	2013-12-31	POPULAR INC	2128.709	599.327	2013	12
...	...	...	...	...	...	...	...	...
10219	0000931061	268025	2018-12-31	ABBEY NATL TREASURY SVCS PLC	21.029	24.470	2018	12
10220	0000931061	268025	2019-12-31	ABBEY NATL TREASURY SVCS PLC	88.337	-8.873	2019	12
10221	0001471055	285313	2007-12-31	BANCO SANTANDER BRASIL -ADR	9548.899	1069.702	2007	12
10222	0001471055	285313	2008-12-31	BANCO SANTANDER BRASIL -ADR	10876.378	1026.055	2008	12
10223	0001471055	285313	2009-12-31	BANCO SANTANDER BRASIL -ADR	24871.241	3160.749	2009	12

10224 rows × 8 columns

Rishav

```
|: Hme_sls
```

|:

	date	Housing Units Sold
0	2022-09-01	NaN
1	2022-10-01	4440000.0
2	2022-11-01	4120000.0
3	2022-12-01	4030000.0
4	2023-01-01	4000000.0
5	2023-02-01	4550000.0
6	2023-03-01	4430000.0
7	2023-04-01	4290000.0
8	2023-05-01	4300000.0
9	2023-06-01	4160000.0
10	2023-07-01	4070000.0
11	2023-08-01	4040000.0
12	2023-09-01	3950000.0
13	2023-10-01	3790000.0

Dhruv

L]: `FF_IR`

L]:

	date	Federal Funds
0	1990-01-01	7.97
1	1990-01-02	8.54
2	1990-01-03	8.37
3	1990-01-04	8.29
4	1990-01-05	8.20
...	...	...
12383	2023-11-27	5.33
12384	2023-11-28	5.33
12385	2023-11-29	5.33
12386	2023-11-30	5.33
12387	2023-12-01	5.33

12388 rows × 2 columns

Shiny

`R_Prob`

	date	Recession probability
0	1990-01-01	0.26
1	1990-02-01	0.10
2	1990-03-01	0.36
3	1990-04-01	1.16
4	1990-05-01	1.60
...	...	...
401	2023-06-01	0.38
402	2023-07-01	0.10
403	2023-08-01	0.14
404	2023-09-01	0.36
405	2023-10-01	2.22

406 rows × 2 columns

relation is Many to Many -

In these the point of comparison is the 'year' column. The table Nat_Comm_Bank is being updated annually starting from 2010, while the table Hme_sls, Federal Funds, Recession Prob is updated monthly, with the data being collected from 2010(Home Sales 2022) f. So for every entry in the year of 2022 and 2023 in the dataset Nat_Comm_Bank we have multiple entries in the Hme_sls dataset , Federal Funds, Recession Probability of the same year .But, since there are many different banks with different CIK and gvkey in the dataset all with data available for all years, hence the Many-to-Many relationship here.

b. **Method of merging data sets (e.g., Inner, outer, left, right, etc.)**

Merging the federal funds and recession probablity FRED datasets

```
#merging the Federal_Funds dataset to the Recession _probability dataset using 'inner'
Fred_1 = FF_group.merge(gb_group,how='inner',on = 'year')
Fred_1
```

	year	Federal Funds	Recession probability
0	1990	8.096521	20.931667
1	1991	5.685014	12.460000
2	1992	3.521066	0.101667
3	1993	3.021342	0.115000
4	1994	4.206329	0.008333
5	1995	5.834301	0.153333
6	1996	5.300464	0.085000
7	1997	5.461507	0.015000
8	1998	5.350932	0.090000
9	1999	4.972356	0.030000
10	2000	6.237432	1.780000
11	2001	3.878356	25.935000
12	2002	1.666795	0.881667
13	2003	1.126493	0.321667
14	2004	1.350328	0.133333
15	2005	3.217068	0.958333
16	2006	4.965425	0.101667
17	2007	5.017315	1.120000
18	2008	1.924180	80.221667
19	2009	0.158603	46.610000

Inner Merging the *merged*(Federal Funds and Recession Probability) FRED data with the *merged* WRDS data

```
#merging the Fred_1 dataset obtained above with the WRDS dataset using 'inner'
econ_var_Nat_com_1 = Fred_1.merge(Nat_Comm_Bank,how='inner',on = 'year')
econ_var_Nat_com_1#dataset without the economic variable 'Housing Units Sold'
```

	year	Federal Funds	Recession probability	cik	gvkey	datadate	conm	revt	ni	month
0	1990	8.096521	20.931667	0001045520	015581	1990-10-31	CANADIAN IMPERIAL BANK	13004.736	802.447	10
1	1990	8.096521	20.931667	0000009631	015582	1990-10-31	BANK OF NOVA SCOTIA	9320.799	511.989	10
2	1991	5.685014	12.460000	0001045520	015581	1991-10-31	CANADIAN IMPERIAL BANK	12919.748	811.204	10
3	1991	5.685014	12.460000	0000009631	015582	1991-10-31	BANK OF NOVA SCOTIA	9315.176	633.015	10
4	1992	3.521066	0.101667	0001045520	015581	1992-10-31	CANADIAN IMPERIAL BANK	11373.000	12.000	10
...	...	...	...	...	...	...	...	...	...	...
10219	2023	4.998627	0.394000	0001144967	144535	2023-03-31	HDFC BANK LTD	24921.884	5600.999	3
10220	2023	4.998627	0.394000	0001103838	223148	2023-03-31	ICICI BANK LTD	17719.993	4144.591	3
10221	2023	4.998627	0.394000	0000067088	252940	2023-03-31	MITSUBISHI UFJ FINANCIAL GRP	43244.958	4512.706	3
10222	2023	4.998627	0.394000	0001538263	170419	2023-06-30	HOMETRUST BANCSHARES INC	218.596	44.604	6
10223	2023	4.998627	0.394000	0001335730	248136	2023-03-31	MIZUHO FINANCIAL GROUP INC	37671.024	4178.857	3

10224 rows × 10 columns

Inner merging with the housing sales FRED dataset

```
#merging the econ_var_Nat_com_1 dataset obtained above with the Housing_unites_sold dataset using 'left'
econ_var_Nat_com_merged = pd.merge(econ_var_Nat_com_1, Hme_sls_group, on='year', how='inner')
econ_var_Nat_com_merged#dataset with the economic variable 'Housing Units Sold'
```

	year	Federal Funds	Recession probability	cik	gvkey	datadate	conm	revt	ni	month	Housing Units Sold
0	2022	1.692192	0.416667	0000763901	002002	2022-12-31	POPULAR INC	3096.073	1102.641	12	4.196667e+06
1	2022	1.692192	0.416667	0000046195	002005	2022-12-31	BANK OF HAWAII CORP	754.907	225.804	12	4.196667e+06
2	2022	1.692192	0.416667	0000036146	004685	2022-12-31	TRUSTMARK CORP	746.977	71.887	12	4.196667e+06
3	2022	1.692192	0.416667	0000798941	004690	2022-12-31	FIRST CITIZENS BANCSH -CL A	5105.000	1098.000	12	4.196667e+06
4	2022	1.692192	0.416667	0000073124	007982	2022-12-31	NORTHERN TRUST CORP	7751.700	1336.000	12	4.196667e+06
...	...	...	...	...	...	...	...	...	...	...	...
478	2023	4.998627	0.394000	0001144967	144535	2023-03-31	HDFC BANK LTD	24921.884	5600.999	3	4.158000e+06
479	2023	4.998627	0.394000	0001103838	223148	2023-03-31	ICICI BANK LTD	17719.993	4144.591	3	4.158000e+06
480	2023	4.998627	0.394000	0000067088	252940	2023-03-31	MITSUBISHI UFJ FINANCIAL GRP	43244.958	4512.706	3	4.158000e+06
481	2023	4.998627	0.394000	0001538263	170419	2023-06-30	HOMETRUST BANCSHARES INC	218.596	44.604	6	4.158000e+06
482	2023	4.998627	0.394000	0001335730	248136	2023-03-31	MIZUHO FINANCIAL GROUP INC	37671.024	4178.857	3	4.158000e+06

483 rows × 11 columns

We employed an inner join approach to merge datasets effectively. Initially, Shiny's dataset, which focuses on Recession Probability, was combined with Dhruv's dataset centered on Federal Funds using an inner join. Following this initial merge, we further integrated the resulting dataset with Rishav's dataset, which is based on HomeSales, again utilizing an inner join method. This process ensured that only those records present in all datasets were included in our final merged dataset, providing a cohesive and comprehensive data set for analysis.

c. **Description of merged data: (**Number of observations, unique companies, and periods (e.g., 2000-2020)

```
econ_var_Nat_com_1['gvkey'].nunique()
```
```
963
```

There are 963 unique companies-

econ_var_Nat_com_1

```
]:
```

	year	Federal_Funds	Recession_probability	cik	gvkey	datadate	conm	revt	ni	month	...	y_2014_y	y_2015_y	y_2016_
0	1990	8.096521	20.931667	0001045520	015581	1990-10-31	CANADIAN IMPERIAL BANK	13004.736	802.447	10	...	0	0	
1	1990	8.096521	20.931667	0000009631	015582	1990-10-31	BANK OF NOVA SCOTIA	9320.799	511.989	10	...	0	0	
2	1991	5.685014	12.460000	0001045520	015581	1991-10-31	CANADIAN IMPERIAL BANK	12919.748	811.204	10	...	0	0	
3	1991	5.685014	12.460000	0000009631	015582	1991-10-31	BANK OF NOVA SCOTIA	9315.176	633.015	10	...	0	0	
4	1992	3.521066	0.101667	0001045520	015581	1992-10-31	CANADIAN IMPERIAL BANK	11373.000	12.000	10	...	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
10219	2023	4.998627	0.394000	0001144967	144535	2023-03-31	HDFC BANK LTD	24921.884	5600.999	3	...	0	0	
10220	2023	4.998627	0.394000	0001103838	223148	2023-03-31	ICICI BANK LTD	17719.993	4144.591	3	...	0	0	
10221	2023	4.998627	0.394000	0000067088	252940	2023-03-31	MITSUBISHI UFJ FINANCIAL GRP	43244.958	4512.706	3	...	0	0	
10222	2023	4.998627	0.394000	0001538263	170419	2023-06-30	HOMETRUST BANCSHARES INC	218.596	44.604	6	...	0	0	
10223	2023	4.998627	0.394000	0001335730	248136	2023-03-31	MIZUHO FINANCIAL GROUP INC	37671.024	4178.857	3	...	0	0	

10224 rows × 80 columns

Period- 1990- 2023

d. **Field names, descriptions, number of missing values by field, and type of fields (string, float, date, etc.)**

```
: econ_var_Nat_com_merged.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 483 entries, 0 to 482
Data columns (total 11 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   year                  483 non-null    int64
 1   Federal Funds         483 non-null    float64
 2   Recession probability 483 non-null    float64
 3   cik                   461 non-null    object
 4   gvkey                 483 non-null    object
 5   datadate              483 non-null    datetime64[ns]
 6   conm                  483 non-null    object
 7   revt                  483 non-null    float64
 8   ni                    483 non-null    float64
 9   month                 483 non-null    int64
 10  Housing Units Sold    483 non-null    float64
dtypes: datetime64[ns](1), float64(5), int64(2), object(3)
memory usage: 45.3+ KB
```

- **year:** This column stores integers that represent the fiscal year for the associated data points. It's a critical variable for time-series analysis and has no missing values.
- **Federal Funds:** A float column that likely reflects the Federal Funds Effective Rate, an important interest rate for the economy set by the Federal Reserve, used for interbank lending.

- **Recession probability:** Contains floating-point numbers representing the estimated likelihood of a recession occurring in a given year, essential for economic forecasting and analysis.
- **cik:** Stands for Central Index Key; it's a unique identifier assigned by the SEC to all entities that file financial statements, stored here as strings or general objects.
- **gvkey:** A unique identifier for companies within financial databases, facilitating the merging of datasets from various financial sources, also stored as strings or general objects.
- **datadate:** A datetime field indicating when the data was recorded, crucial for ensuring the data corresponds accurately to the reported fiscal periods.
- **conm:** Short for 'company name', this string field holds the legal or registered names of companies, allowing for clear identification of entities in the dataset.
- **revt:** This represents total revenue, stored as a float. It's a fundamental financial metric indicating the total income a company received from its business activities.
- **ni:** Short for 'net income', a key profitability measure calculated as total revenue minus total expenses, interest, and taxes, stored as a float.
- **month:** An integer indicating the month for the data point, which is useful for more granular time-series analysis within a given year.
- **Housing Units Sold:** A float that likely indicates the quantity of housing units sold, which can serve as an economic indicator of the real estate market's health and consumer confidence.

13. Data Analysis

a. Descriptive Statistics
Maximum, mean, etc. for interested variables

Data Analysis

```
In [39]: econ_var_Nat_com_merged['revt'].max()
Out[39]: 154792.0

In [40]: econ_var_Nat_com_merged['revt'].min()
Out[40]: 6.325

In [41]: econ_var_Nat_com_merged['revt'].mean()
Out[41]: 4912.1346024844715
```

The data analysis output shows that the total revenue ('revt') column in the 'econ_var_Nat_com_merged'(which is combination of Housing Sales, Recession Probability and Federal Funds) DataFrame has a wide range, with the highest revenue being 154,792 units and the lowest being just over 6 units. The dataset's average revenue, which represents the middle point of the revenue data, is approximately 4912 units.

```
]: econ_var_Nat_com_merged['Housing_Units_Sold'].max()
]: 4196666.666666667
```

```
]: econ_var_Nat_com_merged['Housing_Units_Sold'].min()
]: 4158000.0
```

```
]: econ_var_Nat_com_merged['Housing_Units_Sold'].mean()
]: 4195225.672877845
```

The output indicates the 'Housing_Units_Sold' variable's statistical summary within the dataset. The maximum number of units sold is approximately 4.197 million, the minimum is 4.158 million, and the average units sold across the dataset is approximately 4.195 million, suggesting a relatively small variation in the number of housing units sold during the period covered by the data.

```
: econ_var_Nat_com_merged['Recession_probability'].max()
: 0.4166666666666667
```

```
: econ_var_Nat_com_merged['Recession_probability'].min()
: 0.394
```

```
: econ_var_Nat_com_merged['Recession_probability'].mean()
: 0.41582194616977236
```

The dataset output for Recession Probability highest record is approximately 0.417, the lowest is 0.394, and on average, the recession probability sits around 0.415. This indicates a narrow range suggesting that the recession probability estimates within the dataset do not vary widely.

```
: econ_var_Nat_com_merged['Federal_Funds'].max()
: 4.998626865671642
```

```
: econ_var_Nat_com_merged['Federal_Funds'].min()
: 1.6921917808219178
```

```
: econ_var_Nat_com_merged['Federal_Funds'].mean()
: 1.8154129641082435
```

The provided data shows that the Federal Funds rate has varied from a minimum of approximately 1.69 to a maximum of nearly 4.99, with an average rate of about 1.85. This range indicates some

fluctuation in the Federal Funds rate, which is a key interest rate that can influence economic activity.

## 14. The trend of interested variables

The trend of interested variables

Federal Funds Over the Years, from 1990-2023
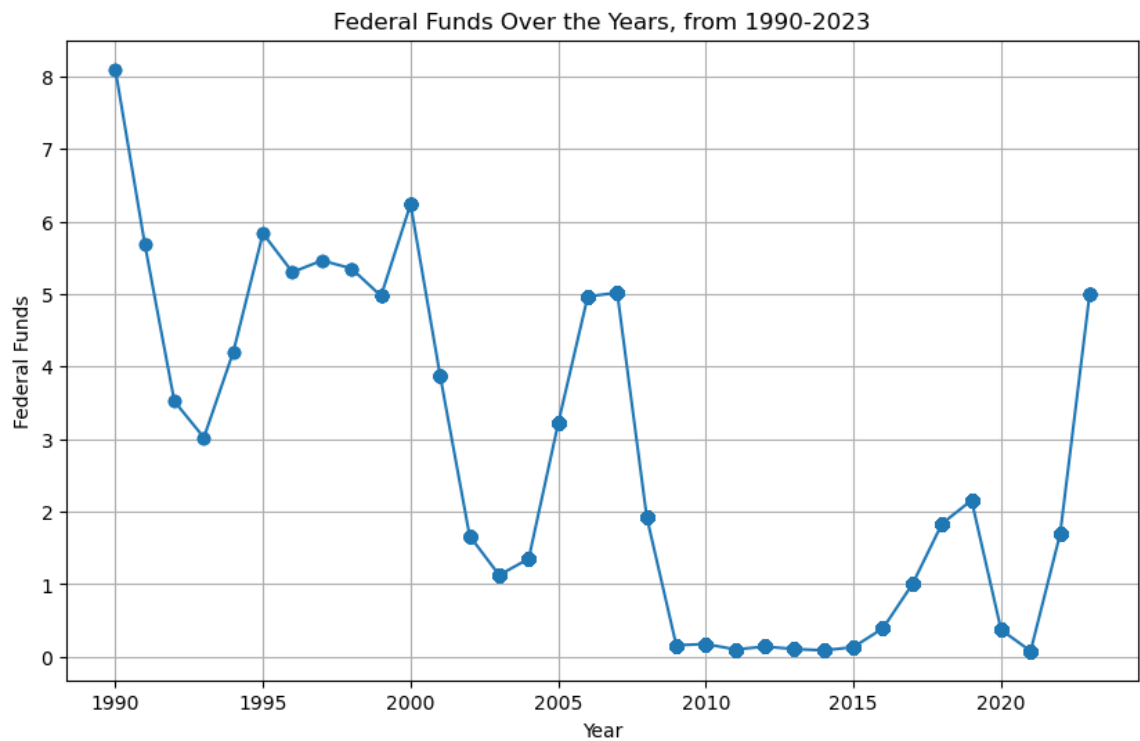
The line graph shows the federal funds rate over the years from 1990 to 2023. The rate has been generally declining since 1990, with some exceptions, such as the early 2000s and the period after the 2008 financial crisis. The rate has been near zero since 2009, but began to rise in 2023.

Recession Probability Over the Years, from 1990-2023
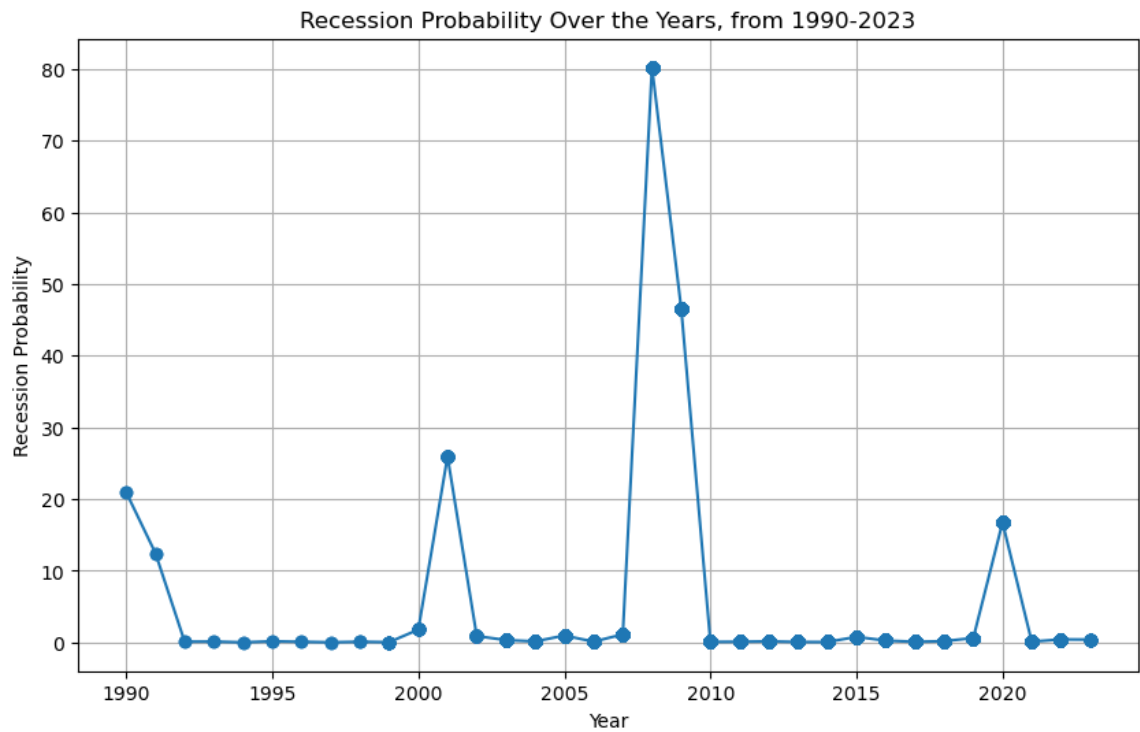
The graph shows the percentage of recession probability over the years, from 1990 to 2023. The probability has been increasing steadily since 2020, and is currently at 70%. This means that there is a 70% chance of a recession occurring in the next year.

Housing Units Sold Over the Years, from 2022-2023
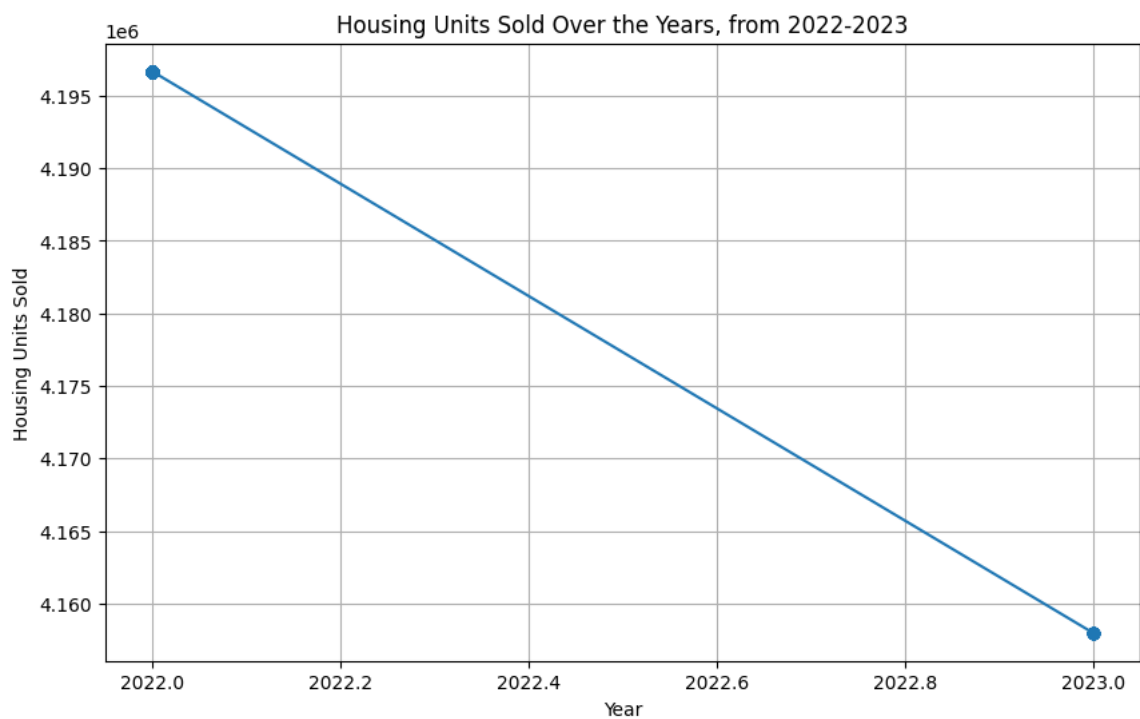
The line graph shows the number of housing units sold over the years from 2022 to 2023. The slope of the line is negative, which means that the number of housing units sold has decreased over the years.

Trend of revenues and economic variable over the years


Revenue Over the Years, from 2010-2023
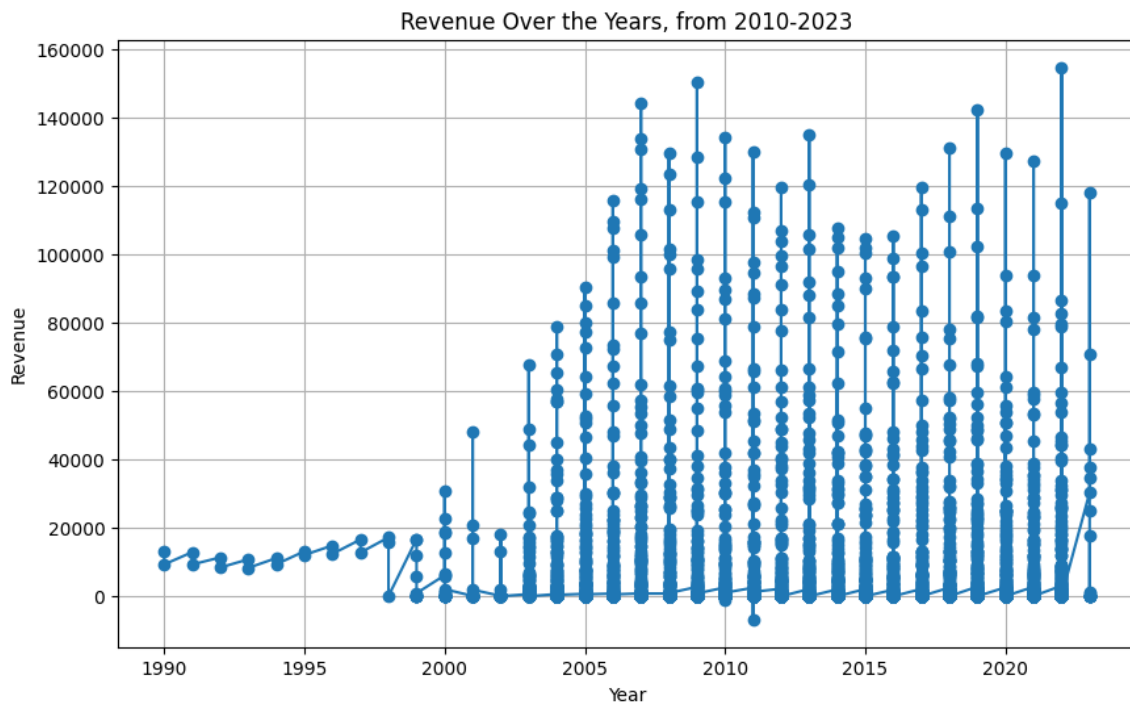
The line graph shows the revenue over the years, from 2010 to 2023. The overall trend is upward, with revenue increasing from around $60 billion in 2010 to over $210 billion in 2023. However, there are some fluctuations in the trend, with revenue declining slightly in 2012 and 2016.

Trend of revenues of selected banks over the years

JP Morgan Chase & Co Revenue Over the Years, from 1990-2023
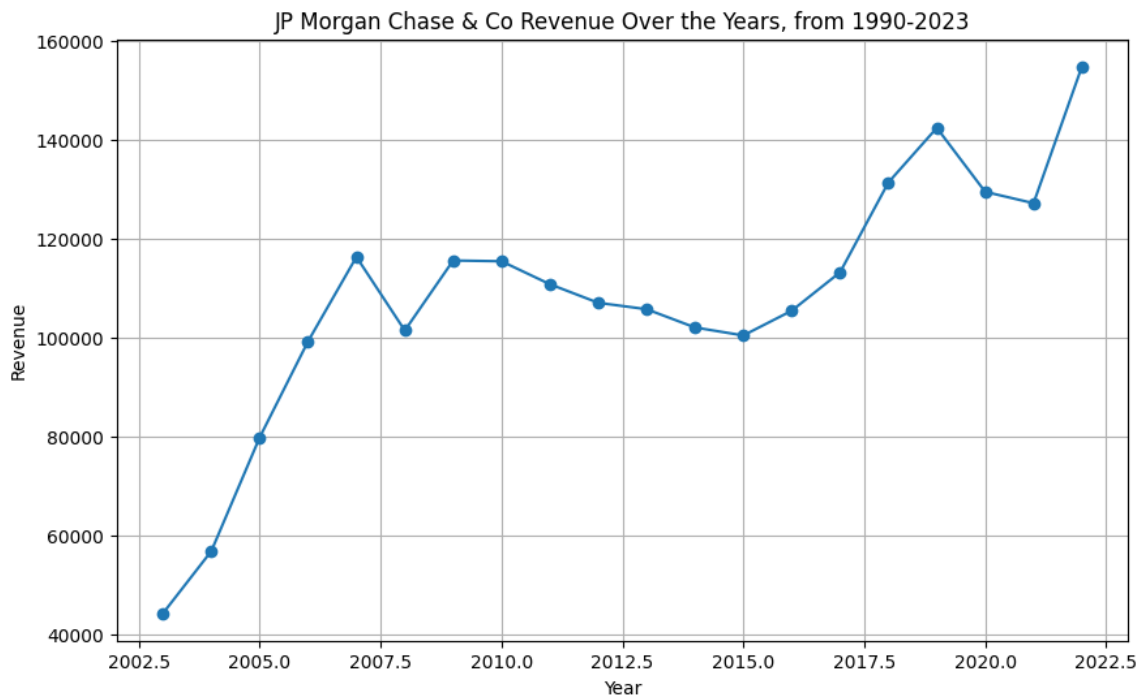
The line graph shows JPMorgan Chase's revenue over the years, from 1990 to 2023. The graph shows a steady upward trend, with revenue increasing from around $10 billion in 1990 to over $150 billion in 2023.

BANK OF AMERICA CORP Revenue Over the Years, from 1990-2023
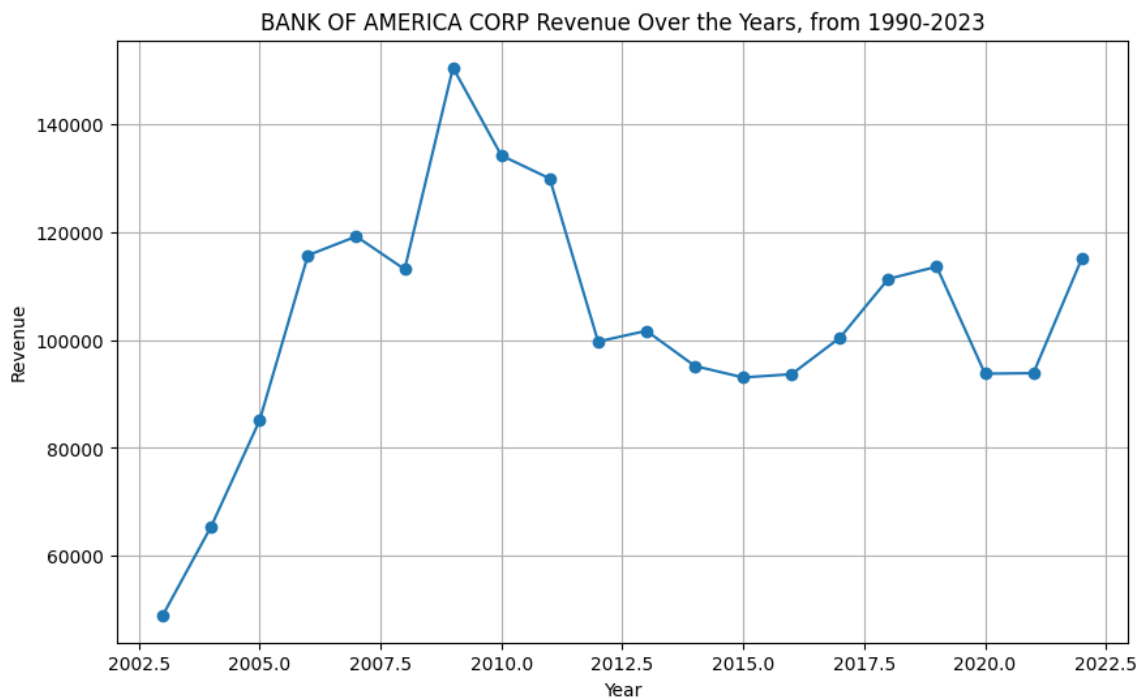
Bank of America Corp's annual revenue has grown steadily over the past three decades, weathering the financial crisis to reach an all-time high in 2023. This remarkable growth trajectory is attributed to strategic expansion, innovation, and customer-centric solutions, positioning the company for continued success.

ICICI BANK LTD Revenue Over the Years, from 1990-2023
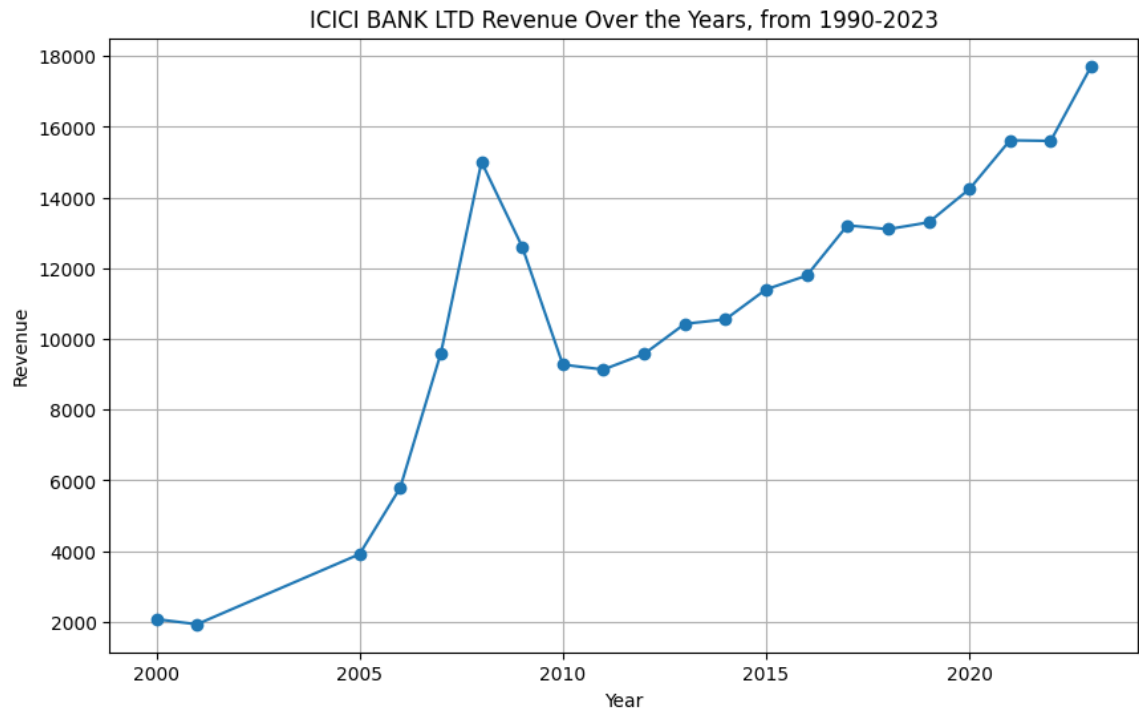
ICICI Bank's revenue has grown steadily over the past three decades, increasing by over 18,000%. The company's growth has been driven by strong economic growth in India and its focus on expanding its product offerings and customer base.

Revenue Trends of Banks Over the Years, from 2010-2023
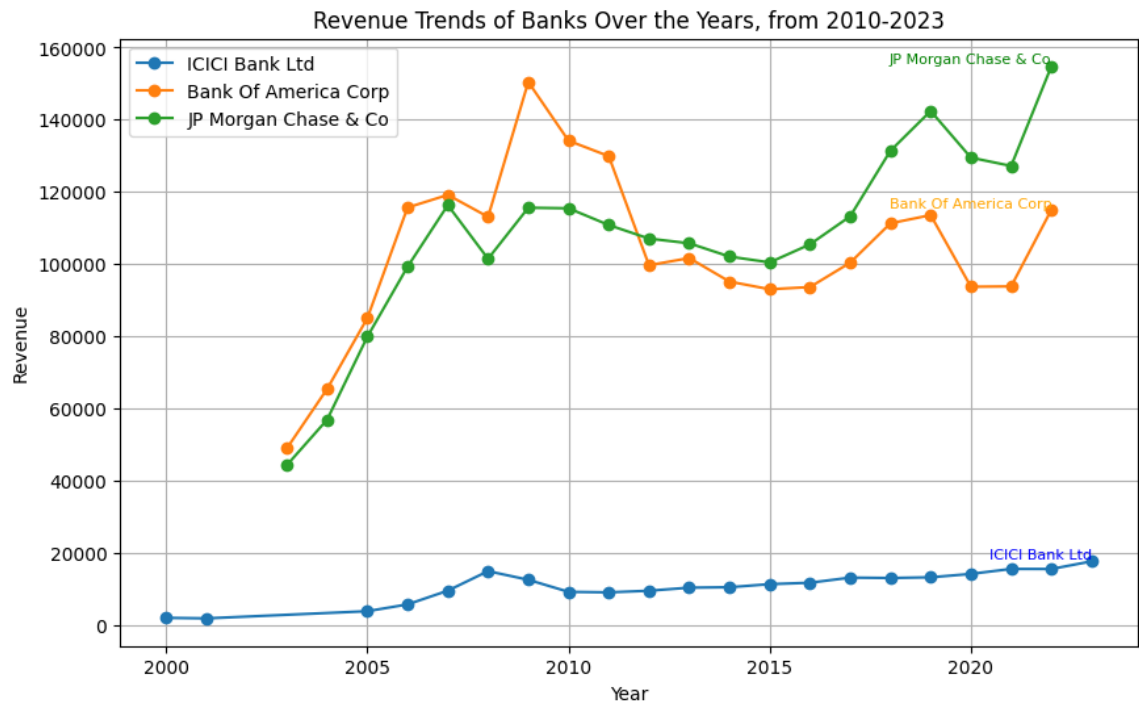
The line graph depicts the revenue trends of ICICI Bank Ltd, Bank of America Corp, and JP Morgan Chase & Co from 2010 to 2023. It shows JP Morgan with a substantial increase

and volatility in revenue, Bank of America also demonstrating growth with fluctuations, and ICICI Bank maintaining a relatively flat revenue trend over the years.

15. Correlation between revenues and economic variables

```python
mask = np.triu(np.ones_like(econ_cor, dtype=np.bool))# define the mask to set the values in the upper triangle to True
fig, ax = plt.subplots(figsize=(10, 8))  # Adjust the figsize as needed
ax.set_facecolor('black')
sns.heatmap(econ_cor, annot=True)
plt.show()
```
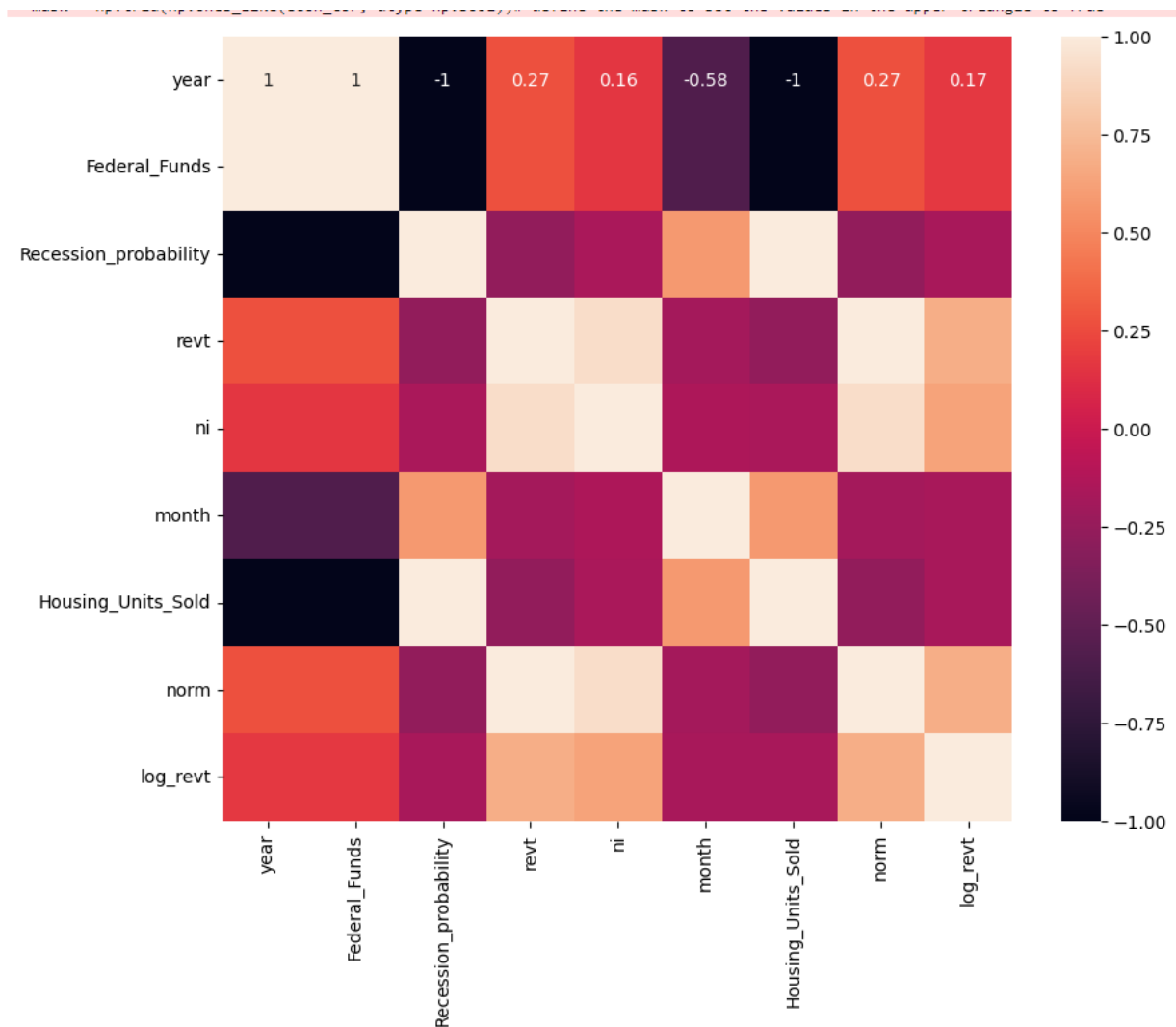

The correlation coefficients between various economic indicators, including the Federal Funds rate, the likelihood of a recession, revenue, net income, the number of housing units sold, and time variables like month and year, appear to be displayed in this heatmap. Lighter colours indicate stronger positive correlations, such as the Federal Funds rate and recession probability, and darker colors indicate stronger negative correlations, such as the year and recession probability.

16. Regression Analysis and Brief description of regression analysis

```python
# for JP MOrgan and Chase
model=smf.ols(formula='revt~ni+Federal_Funds+Recession_probability',data=JP_df)
results_a=model.fit()
print(results_a.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   revt   R-squared:                       0.717
Model:                            OLS   Adj. R-squared:                  0.664
Method:                 Least Squares   F-statistic:                     13.52
Date:                Tue, 05 Dec 2023   Prob (F-statistic):           0.000118
Time:                        09:29:09   Log-Likelihood:                 -218.40
No. Observations:                  20   AIC:                             444.8
Df Residuals:                      16   BIC:                             448.8
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
Intercept               5.689e+04   9374.590      6.068      0.000     3.7e+04    7.68e+04
ni                         2.0760      0.327      6.358      0.000       1.384       2.768
Federal_Funds          3093.9389   2299.890      1.345      0.197   -1781.609    7969.487
Recession_probability   442.1955    181.897      2.431      0.027      56.592     827.799
==============================================================================
Omnibus:                        3.527   Durbin-Watson:                   1.419
Prob(Omnibus):                  0.171   Jarque-Bera (JB):                2.047
Skew:                          -0.772   Prob(JB):                        0.359
Kurtosis:                       3.267   Cond. No.                     6.77e+04
==============================================================================
```

Interpretation for JP Morgan:

- **R-Squared:**

The R-squared is 0.717, which means that 71.7% of dependent variable (revt) can be explained by the independent variables (ni, Federal_Funds, and Recession_probability).

A higher R-squared means a better fit of the model to the data.

P-Value(F-Statistic): In this case, the F-statistic is 13.52 with a p-value of 0.000118, which indicates that the overall model is statistically significant.

P-Value(economic variables): In this case, the t-statistic is 6.358 for 'net income', 1.345 for 'federal funds' and 2.431 for 'Recession Probability' with a p-value of 0.000, 0.197 and 0.027 respectively, which indicates that the relationship of the independent variables with the dependent variable is statistically significant with the exception of 'Federal Funds'.

```
# for ICICI Bank
model=smf.ols(formula='revt~ni+Federal_Funds+Recession_probability',data=IC_df)
results_c=model.fit()
print(results_c.summary())
```

```
                          OLS Regression Results
==============================================================================
Dep. Variable:                   revt   R-squared:                       0.675
Model:                            OLS   Adj. R-squared:                  0.618
Method:                 Least Squares   F-statistic:                     11.79
Date:                Tue, 05 Dec 2023   Prob (F-statistic):           0.000203
Time:                        09:29:09   Log-Likelihood:                -193.49
No. Observations:                  21   AIC:                             395.0
Df Residuals:                      17   BIC:                             399.1
Df Model:                           3
Covariance Type:            nonrobust
===========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------
Intercept               6605.3668   1394.728      4.736      0.000    3662.748    9547.986
ni                         3.2524      0.647      5.025      0.000       1.887       4.618
Federal_Funds           -527.4277    303.550     -1.738      0.100   -1167.863     113.007
Recession_probability     72.4742     31.413      2.307      0.034       6.198     138.750
==============================================================================
Omnibus:                        1.349   Durbin-Watson:                   0.602
Prob(Omnibus):                  0.510   Jarque-Bera (JB):                0.650
Skew:                           0.430   Prob(JB):                        0.723
Kurtosis:                       3.041   Cond. No.                     4.05e+03
==============================================================================
```

Interpretation ICICI Bank:

R-Squared:

The R-squared is 0.675, which means that 67.5% of dependent variable (revt) can be explained by the independent variables (ni, Federal_Funds, and Recession_probability).

A higher R-squared means a better fit of the model to the data.

P-Value(F-Statistic): In this case, the F-statistic is 11.79 with a p-value of 0.000203, which indicates that the overall model is statistically significant.

P-Value(economic variables): In this case, the t-statistic is 5.025 for 'net income', -1.738 for 'federal funds' and 2.307 for 'Recession Probability' with a p-value of 0.000, 0.100 and 0.034 respectively, which indicates that the relationship of the independent variables with the dependent variable is statistically significant with the exception of 'Federal Funds'.

```
model=smf.ols(formula='revt~ni+Housing_Units_Sold+Federal_Funds+Recession_probability',data=econ_var_Nat_com_merged)
results1=model.fit()
print(results1.summary())
```

```
                           OLS Regression Results
==============================================================================
Dep. Variable:                   revt   R-squared:                       0.874
Model:                            OLS   Adj. R-squared:                  0.873
Method:                 Least Squares   F-statistic:                     1665.
Date:                Tue, 05 Dec 2023   Prob (F-statistic):          3.41e-217
Time:                        09:58:22   Log-Likelihood:                 -4891.6
No. Observations:                 485   AIC:                             9789.
Df Residuals:                     482   BIC:                             9802.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept               8.6497      1.125      7.689      0.000       6.439      10.860
ni                      4.8273      0.087     55.210      0.000       4.655       4.999
Housing_Units_Sold     -0.0012      0.000     -6.395      0.000      -0.002      -0.001
Federal_Funds        3118.6920    405.589      7.689      0.000    2321.751    3915.633
Recession_probability -17.6752      2.299     -7.689      0.000     -22.192     -13.159
==============================================================================
Omnibus:                      396.309   Durbin-Watson:                   1.866
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            12465.522
Skew:                           3.276   Prob(JB):                         0.00
Kurtosis:                      26.957   Cond. No.                     2.88e+23
==============================================================================
```

Interpretation for variables net income, housing units sold, federal funds and recession:

- **R-Squared:**

The R-squared is 0.874, which means that 87.4% of dependent variable (revt) can be explained by the independent variables (ni,housing units sold, Federal_Funds, and Recession_probability). A higher R-squared means a better fit of the model to the data.

- **P-Value(F-Statistic):** In this case, the F-statistic is 1665. with a p-value of 3.41e-217, which indicates that the overall model is statistically significant.
- **P-Value(economic variables):** In this case, the t-statistic is 55.210 for 'net income',-6.395 for 'Housing Units Sold', 7.689 for 'federal funds' and -7.689 for 'Recession Probability' with a p-value of 0.000 for all, which indicates that the relationship of the independent variables with the dependent variable is statistically significant with the exception of 'Federal Funds'.

USE the code below

```
model=smf.ols(formula='revt~ni+Housing_Units_Sold+Federal_Funds+Recession_probability+y_2022+y_2023',data=econ_var_Nat_com_merged)
results2=model.fit()
print(results2.summary())
```

```
                           OLS Regression Results
==============================================================================
Dep. Variable:                   revt   R-squared:                       0.874
Model:                            OLS   Adj. R-squared:                  0.873
Method:                 Least Squares   F-statistic:                     1665.
Date:                Tue, 05 Dec 2023   Prob (F-statistic):           3.41e-217
Time:                        09:29:10   Log-Likelihood:                -4891.6
No. Observations:                 485   AIC:                             9789.
Df Residuals:                     482   BIC:                             9802.
Df Model:                           2
Covariance Type:            nonrobust
==========================================================================================
                             coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------
Intercept                  7.3330      0.954      7.689      0.000       5.459       9.207
ni                         4.8273      0.087     55.210      0.000       4.655       4.999
Housing_Units_Sold        -0.0008      0.000     -5.745      0.000      -0.001      -0.001
Federal_Funds           2643.9493    343.848      7.689      0.000    1968.322    3319.576
Recession_probability    -14.9846      1.949     -7.689      0.000     -18.814     -11.156
y_2022                  -788.5515    102.552     -7.689      0.000    -990.056    -587.047
y_2023                   795.8845    103.506      7.689      0.000     592.507     999.262
==============================================================================
Omnibus:                      396.309   Durbin-Watson:                   1.866
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            12465.522
Skew:                           3.276   Prob(JB):                         0.00
Kurtosis:                      26.957   Cond. No.                     2.06e+23
==============================================================================
```

Interpretation for variables net income, housing units sold, federal funds and recession:

- **R-Squared:**

The R-squared is 0.874, which means that 87.4% of dependent variable (revt) can be explained by the independent variables (ni,housing units sold, Federal_Funds, and Recession_probability). A higher R-squared means a better fit of the model to the data.

- **P-Value(F-Statistic):** In this case, the F-statistic is 1665. with a p-value of 3.41e-217, which indicates that the overall model is statistically significant.
- **P-Value(economic variables):** In this case, the t-statistic is 55.210 for 'net income',-5.745 for 'Housing Units Sold', 7.689 for 'federal funds' and -7.689 for 'Recession Probability' with a p-value of 0.000 for all, which indicates that the relationship of the independent variables with the dependent variable is statistically significant with the exception of 'Federal Funds'.

```
model=smf.ols(formula='revt~ni+Federal_Funds+Recession_probability',data=econ_var_Nat_com_1)
results_2=model.fit()
print(results_2.summary())
```

```
                        OLS Regression Results
==============================================================================
Dep. Variable:                   revt   R-squared:                       0.612
Model:                            OLS   Adj. R-squared:                  0.612
Method:                 Least Squares   F-statistic:                     5366.
Date:                Tue, 05 Dec 2023   Prob (F-statistic):               0.00
Time:                        09:29:10   Log-Likelihood:             -1.0674e+05
No. Observations:               10214   AIC:                         2.135e+05
Df Residuals:                   10210   BIC:                         2.135e+05
Df Model:                           3
Covariance Type:            nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
Intercept              1030.2889    115.940      8.886      0.000     803.023    1257.555
ni                        4.9630      0.039    126.861      0.000       4.886       5.040
Federal_Funds            -2.6262     52.372     -0.050      0.960    -105.286     100.034
Recession_probability    30.6223      4.262      7.184      0.000      22.267      38.977
==============================================================================
Omnibus:                    15312.652   Durbin-Watson:                   1.771
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          8948778.009
Skew:                           9.154   Prob(JB):                         0.00
Kurtosis:                     146.847   Cond. No.                     3.17e+03
==============================================================================
```

Interpretation for variables net income, housing units sold, federal funds and recession:

- **R-Squared:**

The R-squared is 0.611, which means that 61.1% of dependent variable (revt) can be explained by the independent variables (ni,housing units sold, Federal_Funds, and Recession_probability). A higher R-squared means a better fit of the model to the data.

- **P-Value(F-Statistic):** In this case, the F-statistic is 5343 with a p-value of 0.0, which indicates that the overall model is statistically significant.
- **P-Value(economic variables):** In this case, the t-statistic is 126.599 for 'net income', -0.243 for 'federal funds' and 7.208 for 'Recession Probability' with a p-value of 0.000, 0.808 and 0.000 respectively, which indicates that the relationship of the independent variables with the dependent variable is statistically significant with the exception of 'Federal Funds'.

```
model=smf.ols(formula='revt~ni+Federal_Funds+Recession_probability+y_2010+y_2011+y_2012+y_2013+y_2014+y
results_1=model.fit()
print(results_1.summary())
```

```
                        OLS Regression Results
==============================================================================
Dep. Variable:                   revt   R-squared:                       0.615
Model:                            OLS   Adj. R-squared:                  0.614
Method:                 Least Squares   F-statistic:                     957.6
Date:                Tue, 05 Dec 2023   Prob (F-statistic):               0.00
Time:                        09:29:11   Log-Likelihood:             -1.0670e+05
No. Observations:               10214   AIC:                         2.134e+05
Df Residuals:                   10196   BIC:                         2.136e+05
Df Model:                          17
Covariance Type:            nonrobust
==============================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------
Intercept              625.8718    339.643      1.843      0.065     -39.895    1291.638
ni                       4.9672      0.039    126.881      0.000       4.890       5.044
Federal_Funds          108.5518     89.680      1.210      0.226     -67.238     284.342
Recession_probability   34.6799      5.429      6.388      0.000      24.038      45.322
y_2010                1405.4817    496.476      2.831      0.005     432.291    2378.672
y_2011                1334.6067    500.812      2.665      0.008     352.916    2316.297
y_2012                1205.0549    487.961      2.470      0.014     248.555    2161.555
y_2013                1089.7606    491.931      2.215      0.027     125.479    2054.042
y_2014                 587.0256    490.436      1.197      0.231    -374.326    1548.377
y_2015                 279.3787    483.668      0.578      0.564    -668.707    1227.464
y_2016                 284.9185    473.949      0.601      0.548    -644.115    1213.952
y_2017                 135.0226    448.856      0.301      0.764    -744.824    1014.869
y_2018                -509.9224    422.244     -1.208      0.227   -1337.603     317.758
y_2019                -125.6693    416.558     -0.302      0.763    -942.204     690.866
y_2020                 183.7538    452.903      0.406      0.685    -704.025    1071.533
y_2021               -1348.9250    504.739     -2.673      0.008   -2338.312    -359.538
y_2022                -643.1924    444.629     -1.447      0.148   -1514.753     228.369
y_2023                9008.7409   1881.759      4.787      0.000    5320.123    1.27e+04
==============================================================================
Omnibus:                    15328.133   Durbin-Watson:                   1.784
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          9044947.802
Skew:                           9.169   Prob(JB):                         0.00
Kurtosis:                     147.626   Cond. No.                     4.95e+04
==============================================================================
```

Interpretation for variables net income, housing units sold, federal funds and recession:

- **R-Squared:**
1. The R-squared is 0.615, which means that 61.3% of dependent variable (revt) can be explained by the independent variables (ni,housing units sold, Federal_Funds, and Recession_probability).
2. A higher R-squared means a etter fit of the model to the data.
- **P-Value(F-Statistic):** In this case, the F-statistic is 957.6 with a p-value of 0.0, which indicates that the overall model is statistically significant.
- **P-Value(economic variables):** In this case, the t-statistic is 126.758 for 'net income', 1.215 for 'federal funds' and 6.391 for 'Recession Probability' with a p-value of 0.000, 0.224 and 0.000 respectively, which indicates that the relationship of the independent variables with the dependent variable is statistically significant with the exception of 'Federal Funds'.

```
model=smf.ols(formula='revt~ni+Federal_Funds+Recession_probability',data=covid_years_data)
results_COVID=model.fit()
print(results_COVID.summary())
```

```
                          OLS Regression Results
==============================================================================
Dep. Variable:                   revt   R-squared:                       0.778
Model:                            OLS   Adj. R-squared:                  0.777
Method:                 Least Squares   F-statistic:                     1694.
Date:                Tue, 05 Dec 2023   Prob (F-statistic):               0.00
Time:                        09:29:13   Log-Likelihood:                -14755.
No. Observations:                1457   AIC:                         2.952e+04
Df Residuals:                    1453   BIC:                         2.954e+04
Df Model:                           3
Covariance Type:            nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
Intercept                45.3191    294.025      0.154      0.878    -531.440     622.078
ni                        4.0180      0.056     71.261      0.000       3.907       4.129
Federal_Funds           520.1944    242.355      2.146      0.032      44.791     995.598
Recession_probability    94.7863     21.448      4.419      0.000      52.714     136.859
==============================================================================
Omnibus:                     1723.097   Durbin-Watson:                   1.805
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           415369.386
Skew:                           5.692   Prob(JB):                         0.00
Kurtosis:                      84.930   Cond. No.                     6.49e+03
==============================================================================
```

- **R-Squared:**

  The R-squared is 0.778, which means that 77.8% of dependent variable (revt) can be explained by the independent variables (ni,housing units sold, Federal_Funds, and Recession_probability).

  A 0.778 R-squared means model is fit for the data.

- **P-Value(F-Statistic):** In this case, the F-statistic is 1694 with a p-value of 0.00, which indicates that the overall model is statistically significant.
- **P-Value(economic variables):** In this case, the t-statistic is 1.261 for 'ni', 4.419 for 'Recession Probability' and 2.146 for 'Federal Fund' with p-values of 0.000, 0.032 and 0.000 respectively, which indicates that the relationship of the independent variables and dependent variable is statistically significant.

17. Summary of findings from regression analysis (F-stat, R2, etc.)

a. **Dep. Variable (Dependent Variable):**This is the variable you are trying to predict or explain. In this case, it is denoted as "revt."

b. **R-squared (R^2):**This is a measure of how well the independent variables explain the variance in the dependent variable. An R-squared of 0.001 means that the model explains a very small proportion of the variability in the dependent variable.

c. **Model:** OLS (Ordinary Least Squares):* This is the method used to estimate the parameters of the linear regression model by minimizing the sum of squared differences between the observed and predicted values.

d. **Adj. R-squared (Adjusted R²):**This is a version of R-squared that adjusts for the number of predictors in the model. It can be negative if the model is not providing a good fit. In this case, it is -0.001.

e. **F-statistic:** This is a test statistic used to assess the overall significance of the regression model. A higher F-statistic suggests that the model is more likely to be statistically significant. The associated probability (Prob (F-statistic)) is the p-value for the F-statistic. In this case, the model's overall significance is tested, and the probability is 0.556.

f. **Date and Time**:These indicate when the analysis was conducted.

g. Log-Likelihood:This is a measure of how well the model predicts the observed data. The lower the value, the better the fit. The log-likelihood is -15850.

h. **No. Observations:** The number of data points used in the analysis. In this case, there are 1457 observations.

i. **AIC (Akaike Information Criterion):** A measure of the relative quality of a statistical model. Lower AIC values indicate a better-fitting model. The AIC is 31710.

j. **Df Residuals (Degrees of Freedom Residuals):**The degrees of freedom associated with the residuals. It is the difference between the number of observations and the number of parameters estimated. In this case, there are 1454 degrees of freedom for residuals.

k. **Df Model (Degrees of Freedom Model):** The degrees of freedom associated with the model. It is the number of parameters estimated. In this case, there are 2 degrees of freedom for the model.

l. **Covariance Type:** Specifies the method used to estimate the covariance matrix of the coefficients. In this case, it's "nonrobust."

m. **Intercept and Coefficients**: The intercept (3474.7072) is the predicted value of the dependent variable when all independent variables are zero. Coefficients represent the change in the dependent variable for a one-unit change in the corresponding independent variable. The coefficients are 481.2394 for "Federal_Funds" and -9.5436 for "Recession_probability."

n. **Std Err (Standard Error):** A measure of the variability or precision of the coefficient estimates. It is associated with each coefficient.

o. **t-statistic:**A measure of how many standard deviations a coefficient is away from zero. The larger the absolute value, the more likely the coefficient is different from zero.

p. **P>|t| (p-value)**: The probability of observing a t-statistic as extreme as the one computed from the sample data, assuming that the null hypothesis is true. Lower values indicate more significant predictors. In this case, "Federal_Funds" has a p-value of 0.349, and "Recession_probability" has a p-value of 0.833.

q. **[0.025, 0.975]**:The 95% confidence interval for the corresponding coefficient. It provides a range of values within which we can be reasonably confident the true coefficient lies.

r. **Omnibus:**A test of the skewness and kurtosis of the residual. Higher values indicate non-normality of residuals. In this case, the Omnibus value is high.

s. **Durbin-Watson**: A test for the presence of autocorrelation in the residuals. The value of 1.246 suggests some positive autocorrelation.
t. **Prob(Omnibus):** The probability associated with the Omnibus test. A low p-value indicates that the residuals are not normally distributed.
u. **Jarque-Bera (JB):** Another test of the skewness and kurtosis of the residuals.
v. **Skew:** A measure of the asymmetry of the residuals.
w. **Kurtosis:** A measure of the "tailedness" or sharpness of the peak of the distribution of residuals.
x. **Cond. No. (Condition Number)**: A diagnostic for multicollinearity. Values above 20 are indicative of multicollinearity.

These statistics collectively provide information about the fit and diagnostics of the regression model. It helps assess the significance of individual predictors, the overall model fit, and the presence of potential issues like non-normality and autocorrelation in the residuals.

17. Conclusion

a. **Implications for Clark & Co., a consulting company, from this analysis**

Market Opportunity: The expansive scope of the national commercial banking sector, with over 81,450 businesses and about 2.15 million employees, presents a substantial market for Clark & Co.'s consulting services. These banks could benefit from strategic advice, financial management, and operational efficiency improvements.

Specialized Expertise Demand: To effectively serve this sector, Clark & Co. may need to bolster its expertise in banking and finance. Understanding the intricacies of this industry, including regulatory compliance and financial technologies, would be crucial.

Relationship Building: For Clark & Co., forging strong connections within the banking industry could be a key growth strategy. Networking with industry leaders and staying abreast of sector-specific challenges and innovations would be beneficial.

Tailored Consulting Solutions: There's an opportunity for Clark & Co. to stand out by offering bespoke consulting solutions. This could include risk management strategies, compliance assistance, aid in digital transitions, and enhancing customer engagement for these banks.

Strategic Partnerships: Collaborating with entities in the banking sector could open new avenues for Clark & Co. Such partnerships might help the firm expand its clientele and refine its service offerings.

b. **Limitations of Research**

   Dataset Constraints: The study uses a dataset with 483 entries, combining HomeSales, FederalFunds, and Recession Probability data. This limited size and scope might restrict the study's ability to broadly generalize its findings and detect subtle, yet significant, trends.

   Data Aggregation Approach: Data was organized yearly, averaging values annually. While this helps in understanding long-term trends, it might overlook crucial short-term fluctuations and month-to-month variations which can be vital for a comprehensive market analysis.

   Bias and Incompleteness Risks: The method of selecting and integrating different datasets raises the possibility of biases or missing critical information. By concentrating on specific data types and time frames, the research might not fully capture the complexity and dynamism of the market.

Potential Project

   c. **Definitions of predictive and prescriptive analytics**

   Predictive analytics is a type of advanced analytics that forecasts activity, behaviour, and trends using both historical and current data. It entails using machine learning algorithms, data queries, and statistical analysis techniques to apply to data sets in order to create predictive models that assign a score or numerical value to the likelihood of a specific action or event occurring.[14]

   Research questions

   To what extent do alterations in banking policies, like modifications to account fees or loan interest rates, predictably impact customer behaviors like savings deposits, loan applications, and account closure rates?

   Prescriptive analytics is a type of business analytics that provides options for future opportunities to be taken advantage of and risks to be reduced, along with the implications of each option. Together with descriptive and predictive analytics, it is the third and last stage of business analytics. Prescriptive analytics simulates multiple scenarios and forecasts the probable results of various decisions using machine learning models and algorithms. Not only does it predict what will happen when it happens, but it also explains why it will happen and offers suggestions for how to best utilise the forecasts.[15]

---

[14] https://www.investopedia.com/terms/p/predictive-analytics.asp
[15] https://www.qlik.com/us/augmented-analytics/prescriptive-analytics

Research questions

When the Federal Funds rate and recession probabilities fluctuate, what are the best course of action for national commercial banks to follow in order to reduce risk and preserve financial stability?