```
# Connect the google drive
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
#Import libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```
# Read CSV as pandas data frame
import pandas as pd
df1=pd.read_csv('/content/drive/MyDrive/CompAndViz/WFHtimeseries_monthly.csv')
df2=pd.read_csv('/content/drive/MyDrive/CompAndViz/before_after_covid.csv')
df3=pd.read_csv('/content/drive/MyDrive/CompAndViz/by_city.csv')
df4=pd.read_csv('/content/drive/MyDrive/CompAndViz/by_industry.csv')
df5=pd.read_csv('/content/drive/MyDrive/CompAndViz/onsite_remote.csv')
df6=pd.read_csv('/content/drive/MyDrive/CompAndViz/Employer_plans.csv')
df7=pd.read_csv('/content/drive/MyDrive/CompAndViz/worker_desire.csv')
```

```
df1
```

	date	WFH_share	fullremote_hist
0	1/1/1965	0.4	NaN
1	1/1/1975	0.6	NaN
2	1/1/1985	4.8	NaN
3	1/1/1993	2.5	NaN
4	1/1/1995	2.2	NaN
...	...	...	...
70	5/1/2024	26.6	NaN
71	6/1/2024	28.6	NaN
72	7/1/2024	29.5	NaN
73	8/1/2024	27.5	NaN
74	9/1/2024	27.7	NaN

75 rows × 3 columns

```
df1['fullremote_hist'].isnull().value_counts()
```

fullremote_hist	count
True	58
False	17

```
df1=df1.drop(columns=['fullremote_hist'])
df1
```

	date	WFH_share
0	1/1/1965	0.4
1	1/1/1975	0.6
2	1/1/1985	4.8
3	1/1/1993	2.5
4	1/1/1995	2.2
...	...	...
70	5/1/2024	26.6
71	6/1/2024	28.6
72	7/1/2024	29.5
73	8/1/2024	27.5
74	9/1/2024	27.7

75 rows x 2 columns

```
df1.isnull()
```

Show hidden output

```
df2
```

	date	wfh	time
0	3/1/2020	4.8	before covid
1	5/1/2020	61.5	covid
2	7/1/2020	51.0	covid
3	8/1/2020	48.3	covid
4	9/1/2020	44.3	covid
5	1/1/2021	48.8	covid
6	2/1/2021	44.5	covid
7	3/1/2021	45.3	covid
8	4/1/2021	45.7	covid
9	5/1/2021	42.6	covid
10	6/1/2021	42.6	covid
11	7/1/2021	43.7	covid
12	8/1/2021	41.3	covid
13	9/1/2021	42.0	covid
14	1/1/2022	40.4	covid
15	2/1/2022	38.6	covid
16	3/1/2022	41.2	covid
17	4/1/2022	39.4	covid
18	5/1/2022	43.6	covid
19	6/1/2022	43.5	covid
20	7/1/2022	41.1	covid
21	8/1/2022	40.1	covid
22	9/1/2022	40.1	covid
23	1/1/2023	36.7	covid
24	2/1/2023	38.2	covid
25	3/1/2023	39.4	covid
26	4/1/2023	40.8	covid
27	7/1/2023	31.8	after covid

```
df3
```

	date	top10_cities_MA	11to50_cities_MA	other_cities_MA6	Atlanta	BayArea	Chicagoland	DC	Dallas	Houston	LosAngeles	Mi
0	1/1/2021	40.1	34.8	32.2	42.8	41.2	31.9	39.3	39.7	37.8	44.2	
1	2/1/2021	40.6	35.4	32.2	43.2	40.5	32.7	40.0	39.4	37.6	43.8	
2	3/1/2021	39.7	34.5	31.2	41.1	39.5	31.5	40.5	38.0	35.7	43.6	
3	4/1/2021	38.7	34.1	30.5	40.2	38.2	31.6	39.5	36.1	33.2	42.1	
4	5/1/2021	38.1	33.4	30.1	39.5	37.2	31.4	39.5	36.1	34.4	41.2	
5	6/1/2021	37.7	33.1	29.7	38.5	39.2	32.0	39.1	36.3	32.3	41.0	
6	7/1/2021	37.8	32.6	29.0	38.4	39.0	34.7	39.4	36.4	31.0	41.6	
7	8/1/2021	37.2	31.8	29.0	37.8	39.6	34.5	38.0	34.5	29.5	41.3	
8	9/1/2021	37.3	31.7	29.0	38.6	42.6	34.3	36.9	32.7	29.9	41.8	
9	1/1/2022	38.6	29.8	28.8	42.7	42.9	35.0	40.3	35.4	35.6	42.3	
10	2/1/2022	38.5	29.6	28.1	43.3	41.9	34.6	39.7	35.1	35.3	42.0	
11	3/1/2022	38.6	29.3	27.8	44.3	40.6	35.8	39.8	36.8	34.4	40.2	
12	4/1/2022	38.5	29.8	27.2	44.2	42.0	36.4	38.7	36.2	35.0	40.0	
13	5/1/2022	37.9	29.6	27.1	42.4	42.1	36.6	37.6	36.0	32.9	40.1	
14	6/1/2022	37.1	30.0	26.9	43.2	40.9	36.3	34.6	35.4	32.6	38.1	
15	7/1/2022	36.1	30.3	26.8	42.9	38.6	33.9	33.1	33.4	29.3	35.8	
16	8/1/2022	35.8	30.4	26.6	40.7	37.3	32.9	33.4	34.4	29.5	36.0	
17	9/1/2022	34.5	30.5	26.5	38.8	35.9	31.5	33.6	31.9	28.1	36.4	
18	1/1/2023	33.3	29.0	25.0	33.7	33.0	29.6	34.0	32.1	33.3	35.9	
19	2/1/2023	33.1	28.8	24.8	33.5	34.2	30.6	31.9	32.5	34.6	35.2	
20	3/1/2023	33.5	27.6	25.0	32.1	33.8	31.3	30.9	34.0	36.2	34.7	
21	4/1/2023	33.6	27.0	24.7	33.8	32.5	31.4	32.9	34.6	34.7	34.3	
22	5/1/2023	34.4	27.4	25.0	33.8	34.9	32.4	31.7	36.2	34.2	34.9	
23	6/1/2023	34.3	27.6	26.2	32.8	36.8	31.8	30.5	35.3	34.6	35.1	
24	7/1/2023	34.5	28.9	27.8	32.4	37.6	31.4	29.1	35.1	34.1	35.4	
25	8/1/2023	34.6	29.2	27.8	32.4	39.4	28.8	33.1	33.8	31.9	36.3	
26	1/1/2024	32.2	28.1	26.8	29.0	36.8	30.2	33.3	30.6	29.9	38.5	
27	2/1/2024	31.7	27.6	26.9	28.9	33.5	31.7	31.6	32.6	30.5	36.7	
28	3/1/2024	31.3	27.6	26.5	28.8	32.1	30.0	31.4	32.7	32.0	35.0	
29	4/1/2024	31.1	27.6	26.4	30.9	30.4	29.5	29.9	34.4	30.5	34.7	
30	5/1/2024	31.1	27.8	26.0	31.4	31.4	29.8	31.4	33.1	31.3	33.9	
31	6/1/2024	31.0	27.6	26.1	31.8	31.4	30.4	31.3	32.2	29.2	32.7	
32	7/1/2024	31.4	27.6	25.9	31.8	30.9	31.9	32.9	34.8	28.5	32.0	
33	8/1/2024	31.5	27.8	26.1	32.0	32.4	32.5	33.4	34.1	28.7	32.6	
34	9/1/2024	31.6	28.5	26.4	32.3	31.1	34.4	34.1	35.4	28.1	32.5	

```
df3.columns
```

```
Index(['date', 'top10_cities_MA', '11to50_cities_MA', 'other_cities_MA6',
       'Atlanta', 'BayArea', 'Chicagoland', 'DC', 'Dallas', 'Houston',
       'LosAngeles', 'Miami', 'NewYork'],
      dtype='object')
```

```
# Convert from wide to long format
df_city = pd.melt(df3, id_vars=['date'], value_vars=['top10_cities_MA', '11to50_cities_MA', 'other_cities_MA6','Atlanta','BayArea','Chicagol
                  var_name='City', value_name='WFH')

# Display the result
print(df_city)
```

```
         date             City   WFH
0    1/1/2021  top10_cities_MA  40.1
1    2/1/2021  top10_cities_MA  40.6
2    3/1/2021  top10_cities_MA  39.7
3    4/1/2021  top10_cities_MA  38.7
4    5/1/2021  top10_cities_MA  38.1
..        ...              ...   ...
415  5/1/2024          NewYork  31.2
416  6/1/2024          NewYork  31.5
417  7/1/2024          NewYork  32.1
418  8/1/2024          NewYork  31.5
419  9/1/2024          NewYork  31.2

[420 rows x 3 columns]
```

```
df_city
```

	date	City	WFH
0	1/1/2021	top10_cities_MA	40.1
1	2/1/2021	top10_cities_MA	40.6
2	3/1/2021	top10_cities_MA	39.7
3	4/1/2021	top10_cities_MA	38.7
4	5/1/2021	top10_cities_MA	38.1
...	...	...	...
415	5/1/2024	NewYork	31.2
416	6/1/2024	NewYork	31.5
417	7/1/2024	NewYork	32.1
418	8/1/2024	NewYork	31.5
419	9/1/2024	NewYork	31.2

420 rows x 3 columns

```
df_city.to_csv('CityWfh.csv')
```

```
df5
```

	date	full_onsite	ybrid	full_remote
0	11/1/2021	54.5	30.3	15.2
1	12/1/2021	53.5	32.5	14.0
2	1/1/2022	56.8	25.4	17.7
3	2/1/2022	59.5	22.8	17.7
4	3/1/2022	57.3	27.2	15.5
5	4/1/2022	59.2	27.8	13.0
6	5/1/2022	55.6	30.7	13.7
7	6/1/2022	53.0	33.4	13.6
8	7/1/2022	57.4	28.4	14.2
9	8/1/2022	56.3	29.4	14.3
10	9/1/2022	57.0	30.2	12.9
11	10/1/2022	57.2	29.7	13.1
12	11/1/2022	59.8	27.4	12.9
13	12/1/2022	57.5	29.5	13.0
14	1/1/2023	61.8	26.3	11.9
15	2/1/2023	60.7	28.3	11.0
16	3/1/2023	58.2	29.1	12.7
17	4/1/2023	57.7	31.6	10.7
18	5/1/2023	60.3	27.5	12.2
19	6/1/2023	58.3	29.3	12.4
20	7/1/2023	57.8	28.1	14.1
21	8/1/2023	55.8	32.3	11.9
22	9/1/2023	57.8	30.7	11.5
23	10/1/2023	59.8	29.0	11.2
24	11/1/2023	58.2	28.8	13.1
25	12/1/2023	57.1	30.2	12.7
26	1/1/2024	57.8	29.6	12.6
27	2/1/2024	61.6	24.7	13.7
28	3/1/2024	61.3	26.3	12.4
29	4/1/2024	61.2	27.2	11.6
30	5/1/2024	64.0	23.7	12.3
31	6/1/2024	57.9	29.9	12.1
32	7/1/2024	59.4	25.7	14.9
33	8/1/2024	62.7	23.1	14.2
34	9/1/2024	61.6	25.1	13.3

```
df5 = df5.rename(columns={'ybrid':'hybrid'})

df5
```

	date	full_onsite	hybrid	full_remote
0	11/1/2021	54.5	30.3	15.2
1	12/1/2021	53.5	32.5	14.0
2	1/1/2022	56.8	25.4	17.7
3	2/1/2022	59.5	22.8	17.7
4	3/1/2022	57.3	27.2	15.5
5	4/1/2022	59.2	27.8	13.0
6	5/1/2022	55.6	30.7	13.7
7	6/1/2022	53.0	33.4	13.6
8	7/1/2022	57.4	28.4	14.2
9	8/1/2022	56.3	29.4	14.3
10	9/1/2022	57.0	30.2	12.9
11	10/1/2022	57.2	29.7	13.1
12	11/1/2022	59.8	27.4	12.9
13	12/1/2022	57.5	29.5	13.0
14	1/1/2023	61.8	26.3	11.9
15	2/1/2023	60.7	28.3	11.0
16	3/1/2023	58.2	29.1	12.7
17	4/1/2023	57.7	31.6	10.7
18	5/1/2023	60.3	27.5	12.2
19	6/1/2023	58.3	29.3	12.4
20	7/1/2023	57.8	28.1	14.1
21	8/1/2023	55.8	32.3	11.9
22	9/1/2023	57.8	30.7	11.5
23	10/1/2023	59.8	29.0	11.2
24	11/1/2023	58.2	28.8	13.1
25	12/1/2023	57.1	30.2	12.7
26	1/1/2024	57.8	29.6	12.6
27	2/1/2024	61.6	24.7	13.7
28	3/1/2024	61.3	26.3	12.4
29	4/1/2024	61.2	27.2	11.6
30	5/1/2024	64.0	23.7	12.3
31	6/1/2024	57.9	29.9	12.1
32	7/1/2024	59.4	25.7	14.9
33	8/1/2024	62.7	23.1	14.2
34	9/1/2024	61.6	25.1	13.3

```
df5.columns
```

```
Index(['date', 'full_onsite', 'hybrid', 'full_remote'], dtype='object')
```

```
# Convert from wide to long format
df_workmode = pd.melt(df5, id_vars=['date'], value_vars=['full_onsite', 'hybrid', 'full_remote'], var_name='Work_mode', value_name='WFH')

# Display the result
print(df_workmode)
```

```
         date    Work_mode   WFH
0    11/1/2021  full_onsite  54.5
1    12/1/2021  full_onsite  53.5
2     1/1/2022  full_onsite  56.8
3     2/1/2022  full_onsite  59.5
```

```
4     3/1/2022  full_onsite  57.3
..         ...          ...   ...
100   5/1/2024  full_remote  12.3
101   6/1/2024  full_remote  12.1
102   7/1/2024  full_remote  14.9
103   8/1/2024  full_remote  14.2
104   9/1/2024  full_remote  13.3

[105 rows x 3 columns]
```

```
df_workmode
```

	date	Work_mode	WFH
0	11/1/2021	full_onsite	54.5
1	12/1/2021	full_onsite	53.5
2	1/1/2022	full_onsite	56.8
3	2/1/2022	full_onsite	59.5
4	3/1/2022	full_onsite	57.3
...	...	...	...
100	5/1/2024	full_remote	12.3
101	6/1/2024	full_remote	12.1
102	7/1/2024	full_remote	14.9
103	8/1/2024	full_remote	14.2
104	9/1/2024	full_remote	13.3

105 rows x 3 columns

```
df_workmode.to_csv('Work_mode.csv')
```

```
df4
```

	date	full_onsite-arts_entertain	full_onsite-education	full_onsite-finance_insurance	full_onsite-government	full_onsite-healthcare	full_onsite-hospitaility_food	full_onsite-information	full_onsite-manufacturing
0	1/1/2022	31.4	61.3	25.7	62.2	56.9	78.6	29.5	71.2
1	2/1/2022	41.0	61.9	28.7	61.0	60.9	79.8	24.1	72.8
2	3/1/2022	36.7	64.7	29.2	62.4	60.6	79.2	22.2	74.4
3	4/1/2022	38.6	64.1	28.0	65.1	60.5	79.6	20.3	73.5
4	5/1/2022	43.6	58.2	27.9	61.0	57.9	80.2	22.6	73.1
5	6/1/2022	48.7	55.7	28.1	58.5	58.1	78.9	24.0	70.0
6	7/1/2022	43.4	53.3	29.5	53.4	58.1	78.5	23.4	70.0
7	8/1/2022	42.5	57.6	30.8	58.7	57.4	82.2	23.1	67.5
8	9/1/2022	46.6	59.3	31.2	57.2	58.1	83.0	23.2	67.5
9	1/1/2023	51.2	68.1	35.4	62.3	57.5	78.1	22.0	72.2
10	2/1/2023	45.2	67.2	33.8	61.4	58.6	82.0	22.5	70.6
11	3/1/2023	41.7	66.2	31.2	60.2	58.2	81.7	23.3	69.3
12	4/1/2023	48.7	65.8	29.6	58.2	57.3	82.1	25.5	68.4
13	5/1/2023	49.0	64.6	30.2	61.5	58.1	81.4	26.3	67.1
14	6/1/2023	45.6	62.4	32.4	61.5	59.6	81.1	24.4	68.1
15	7/1/2023	35.9	59.9	30.6	64.9	60.3	80.9	26.6	67.9
16	8/1/2023	33.0	59.2	31.5	67.0	60.9	78.6	24.3	72.3
17	1/1/2024	40.8	69.0	29.2	61.0	60.8	76.5	29.2	71.5
18	2/1/2024	41.7	70.3	32.5	60.7	61.6	78.4	28.5	73.9
19	3/1/2024	39.2	69.6	33.1	58.1	62.9	82.2	31.0	73.2
20	4/1/2024	41.5	67.0	35.5	60.5	61.1	83.5	32.2	71.4
21	5/1/2024	43.0	65.0	31.9	62.5	59.0	81.2	31.7	70.0
22	6/1/2024	46.3	64.3	29.9	60.5	57.6	81.6	31.2	70.1
23	7/1/2024	45.6	65.7	29.9	55.9	60.1	79.2	30.7	69.2
24	8/1/2024	50.0	69.4	31.5	57.0	60.5	82.1	29.8	70.0
25	9/1/2024	49.6	74.0	33.6	58.8	61.9	81.0	28.4	70.5

26 rows × 43 columns

```
df_industry = pd.melt(df4, id_vars=['date'], var_name='Work_Mode - Industry', value_name='Wfh')

print("Before splitting:\n", df_industry)
```

```
Before splitting:
            date        Work_Mode - Industry   Wfh
0     1/1/2022  full_onsite-arts_entertain  31.4
1     2/1/2022  full_onsite-arts_entertain  41.0
2     3/1/2022  full_onsite-arts_entertain  36.7
3     4/1/2022  full_onsite-arts_entertain  38.6
4     5/1/2022  full_onsite-arts_entertain  43.6
...        ...                         ...   ...
1087  5/1/2024       full_remote-wholesale  13.2
1088  6/1/2024       full_remote-wholesale  17.8
1089  7/1/2024       full_remote-wholesale  14.5
1090  8/1/2024       full_remote-wholesale  19.7
1091  9/1/2024       full_remote-wholesale  18.0

[1092 rows x 3 columns]
```

```
df_industry
```

	date	Work_Mode - Industry	Wfh
0	1/1/2022	full_onsite-arts_entertain	31.4
1	2/1/2022	full_onsite-arts_entertain	41.0
2	3/1/2022	full_onsite-arts_entertain	36.7
3	4/1/2022	full_onsite-arts_entertain	38.6
4	5/1/2022	full_onsite-arts_entertain	43.6
...	...	...	...
1087	5/1/2024	full_remote-wholesale	13.2
1088	6/1/2024	full_remote-wholesale	17.8
1089	7/1/2024	full_remote-wholesale	14.5
1090	8/1/2024	full_remote-wholesale	19.7
1091	9/1/2024	full_remote-wholesale	18.0

1092 rows × 3 columns

```
#Splitting the work-mode and Industry as two columns
df_industry[['Work_Mode', 'Industry']] = df_industry['Work_Mode - Industry'].str.split('-', expand=True)

# Drop the original 'Work_Mode - Industry' column
df = df_industry.drop(columns=['Work_Mode - Industry'])
```

```
df
```

	date	Wfh	Work_Mode	Industry
0	1/1/2022	31.4	full_onsite	arts_entertain
1	2/1/2022	41.0	full_onsite	arts_entertain
2	3/1/2022	36.7	full_onsite	arts_entertain
3	4/1/2022	38.6	full_onsite	arts_entertain
4	5/1/2022	43.6	full_onsite	arts_entertain
...	...	...	...	...
1087	5/1/2024	13.2	full_remote	wholesale
1088	6/1/2024	17.8	full_remote	wholesale
1089	7/1/2024	14.5	full_remote	wholesale
1090	8/1/2024	19.7	full_remote	wholesale
1091	9/1/2024	18.0	full_remote	wholesale

1092 rows × 4 columns

```
df.to_csv('industrywfh.csv')
```

```
df6
```