IMPORT IMPORTANT LIBRARIES

```
!pip install openpyxl --quiet

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

plt.rcParams['figure.figsize'] = (10, 5)
plt.rcParams['axes.grid'] = True
plt.rcParams['font.size'] = 11
```

```
from google.colab import files

uploaded = files.upload()
list(uploaded.keys())
```

Choose Files WFHtimese…onthly.xlsx

WFHtimeseries_monthly.xlsx(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 89851 bytes, last modified: 11/19/2025 - 100% done

```
Saving WFHtimeseries_monthly.xlsx to WFHtimeseries_monthly (1).xlsx
['WFHtimeseries_monthly (1).xlsx']
```

```
excel_file = 'WFHtimeseries_monthly.xlsx'

all_sheets = pd.read_excel(excel_file, sheet_name=None)

print("Sheets in this workbook:")
for name in all_sheets.keys():
    print("-", name)
```

```
Sheets in this workbook:
- README - dictionary & content
- WFH before-during COVID
- WFH 1965 - present
- Employer Plans post-COVID WFH
- Worker Desires post-COVID WFH
- Full Remote-Hybrid-Full Onsite
- Workday-weighted WFH series
- WFH by city
- Work Arrangements by Industry
- WFH Rates by Industry
- WFH Rates for Women & Men
- LEGACY WFH series
```

```
import pandas as pd

excel_file = "WFHtimeseries_monthly.xlsx"

all_sheets = pd.read_excel(excel_file, sheet_name=None)

all_sheets.keys()
```

```
dict_keys(['README - dictionary & content', 'WFH before-during COVID',
'WFH 1965 - present', 'Employer Plans post-COVID WFH', 'Worker Desires
post-COVID WFH', 'Full Remote-Hybrid-Full Onsite', 'Workday-weighted WFH
series', 'WFH by city', 'Work Arrangements by Industry', 'WFH Rates by
Industry', 'WFH Rates for Women & Men', 'LEGACY WFH series'])
```

```
df = all_sheets["WFH before-during COVID"]
```

```
df.head()
```

	date	wfhcovid_matquestion	wfhcovid_frac_HPS	Notes	License	Citat
0	2020-03-01	7.152462	NaN	Pre-COVID value is Authors' estimate using dat...	NaN	N
1	2020-05-01	61.563080	NaN	NaN	Copyright 2025 by Jose Maria Barrero,	W using w ple (Barr

Next steps: Generate code with df New interactive sheet

CREATING DATAFRAMES FOR THE KEY SHEETS

```
df_before_during = all_sheets['WFH before-during COVID']
df_emp_plans     = all_sheets['Employer Plans post-COVID WFH']
df_worker_desire = all_sheets['Worker Desires post-COVID WFH']
df_modes         = all_sheets['Full Remote-Hybrid-Full Onsite']
df_city          = all_sheets['WFH by city']
df_industry      = all_sheets['Work Arrangements by Industry']
```

```
print("=== WFH before-during COVID ===")
display(df_before_during.head())

print("=== Employer Plans ===")
display(df_emp_plans.head())

print("=== Worker Desires ===")
display(df_worker_desire.head())

print("=== Full Remote/Hybrid/Onsite ===")
display(df_modes.head())

print("=== WFH by city ===")
display(df_city.head())

print("=== Work Arrangements by Industry ===")
display(df_industry.head())
```

```
=== WFH before-during COVID ===
```

	date	wfhcovid_matquestion	wfhcovid_frac_HPS	Notes	License	Citat
0	2020-03-01	7.152462	NaN	Pre-COVID value is Authors' estimate using dat...	NaN	N
1	2020-05-01	61.563080	NaN	NaN	Copyright 2025 by Jose Maria Barrero, Nicholas...	W using w ple c Barr
2	2020-06-01	56.369545	NaN	The SWAA June 2020 estimate is averages the Ma...	NaN	N
3	2020-07-01	51.176006	NaN	NaN	NaN	N
4	2020-08-01	48.404514	NaN	NaN	NaN	N

```
=== Employer Plans ===
```

	date	wfh_days_postCOVID_planMAd	wfh_days_postCOVID_plan_eMAd	wfhcov
0	2020-05-01	NaN	NaN	
1	2020-07-01	1.057038	NaN	
2	2020-08-01	1.059582	1.577442	
3	2020-09-01	1.091445	1.558076	
4	2020-10-01	1.136645	1.577953	

```
=== Worker Desires ===
```

	date	wfh_days_postCOVID_desMAd	wfh_days_postCOVID_des_eMAd	wfhcovid_
0	2020-05-01	2.090485	NaN	
1	2020-07-01	2.099195	NaN	
2	2020-08-01	2.168038	2.582880	
3	2020-09-01	2.197284	2.547624	
4	2020-10-01	2.314158	2.637780	

```
=== Full Remote/Hybrid/Onsite ===
```

	date	full_onsite_curr	hybrid_curr	full_remote_curr	full_onsite_curr
0	2021-11-01	54.361965	30.379524	15.258510	30.897
1	2021-12-01	53.439350	32.561432	13.999217	29.128
2	2022-01-01	56.790363	25.459003	17.750633	32.101
3	2022-02-01	59.484886	22.782024	17.733088	31.127
4	2022-03-01	57.281601	27.246471	15.471930	30.511

```
=== WFH by city ===
```

	date	wfhcovid_series_top10_MA6_	wfhcovid_series_11to50_MA6_	wfhcovi
0	2020-10-01	51.067982	44.005394	
1	2020-	42.185829	37.142673	

11-01

CLEAN THE DATE COLUMN

```
import pandas as pd

def yyyymm_to_datetime(series):
    """
    Convert an integer/string YYYYMM series to a pandas datetime (first
    """
    s = series.astype(str)

    if not s.str.fullmatch(r'\d{6}').all():

        return pd.to_datetime(series, errors='coerce')

    year = s.str.slice(0, 4).astype(int)
    month = s.str.slice(4, 6).astype(int)
    return pd.to_datetime(dict(year=year, month=month, day=1))

for df_item in [df_before_during, df_emp_plans, df_worker_desire, df_mod
    if 'date' in df_item.columns:

        if not pd.api.types.is_datetime64_any_dtype(df_item['date']):
            df_item['date'] = yyyymm_to_datetime(df_item['date'])

df_before_during[['date']].head()
```

```
         date
0  2020-03-01
1  2020-05-01
2  2020-06-01
3  2020-07-01
4  2020-08-01
```

```
import pandas as pd

excel_file = "WFHtimeseries_monthly.xlsx"

all_sheets = pd.read_excel(excel_file, sheet_name=None)

print("Sheets in workbook:")
for name in all_sheets.keys():
    print("-", name)
```

```
Sheets in workbook:
- README - dictionary & content
- WFH before-during COVID
- WFH 1965 - present
- Employer Plans post-COVID WFH
- Worker Desires post-COVID WFH
- Full Remote-Hybrid-Full Onsite
- Workday-weighted WFH series
- WFH by city
- Work Arrangements by Industry
- WFH Rates by Industry
- WFH Rates for Women & Men
- LEGACY WFH series
```

```
df_before_during = all_sheets["WFH before-during COVID"]

df_emp_plans = all_sheets["Employer Plans post-COVID WFH"]

df_worker_desire = all_sheets["Worker Desires post-COVID WFH"]

df_modes = all_sheets["Full Remote-Hybrid-Full Onsite"]

df_city = all_sheets["WFH by city"]

df_industry = all_sheets["Work Arrangements by Industry"]
```

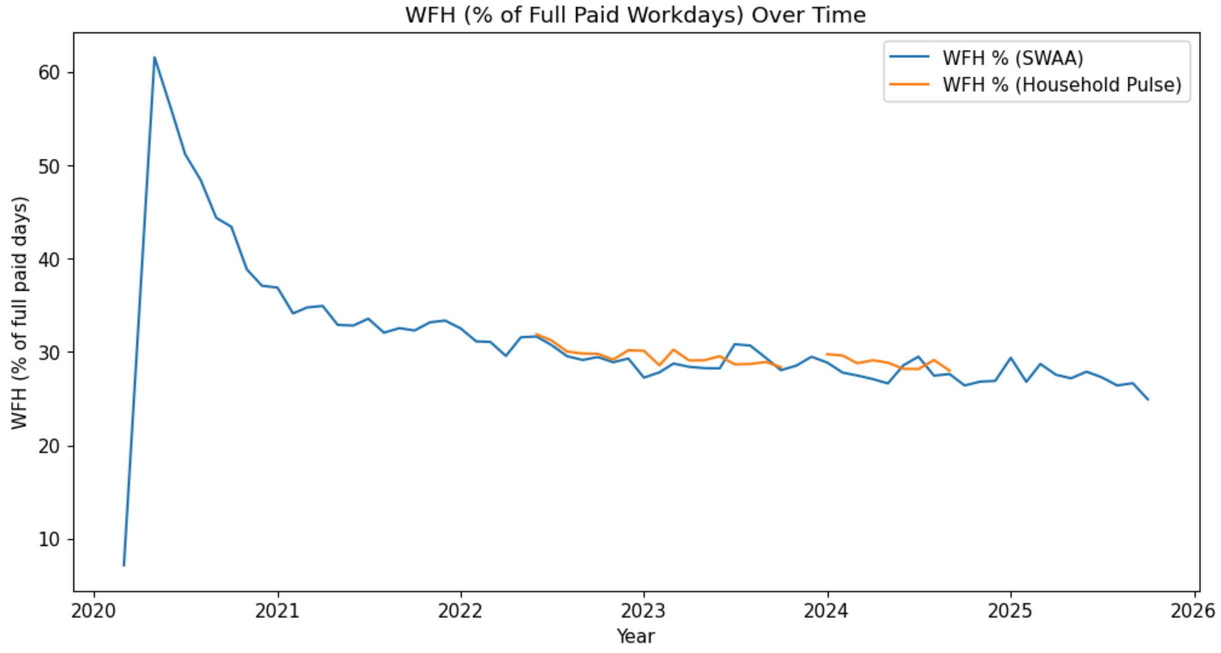
```
df_before_during.head()
df_emp_plans.head()
df_worker_desire.head()
df_modes.head()
df_city.head()
df_industry.head()
```

	date	full_onsite_arts_entertain	full_onsite_education	full_onsite_f
0	2021-11-01	21.706787	64.135490	
1	2021-12-01	29.062971	61.601421	
2	2022-01-01	31.446100	61.309658	
3	2022-02-01	40.977402	61.947681	
4	2022-03-01	36.836796	64.806114	

5 rows × 46 columns

WFH Before / During / After COVID

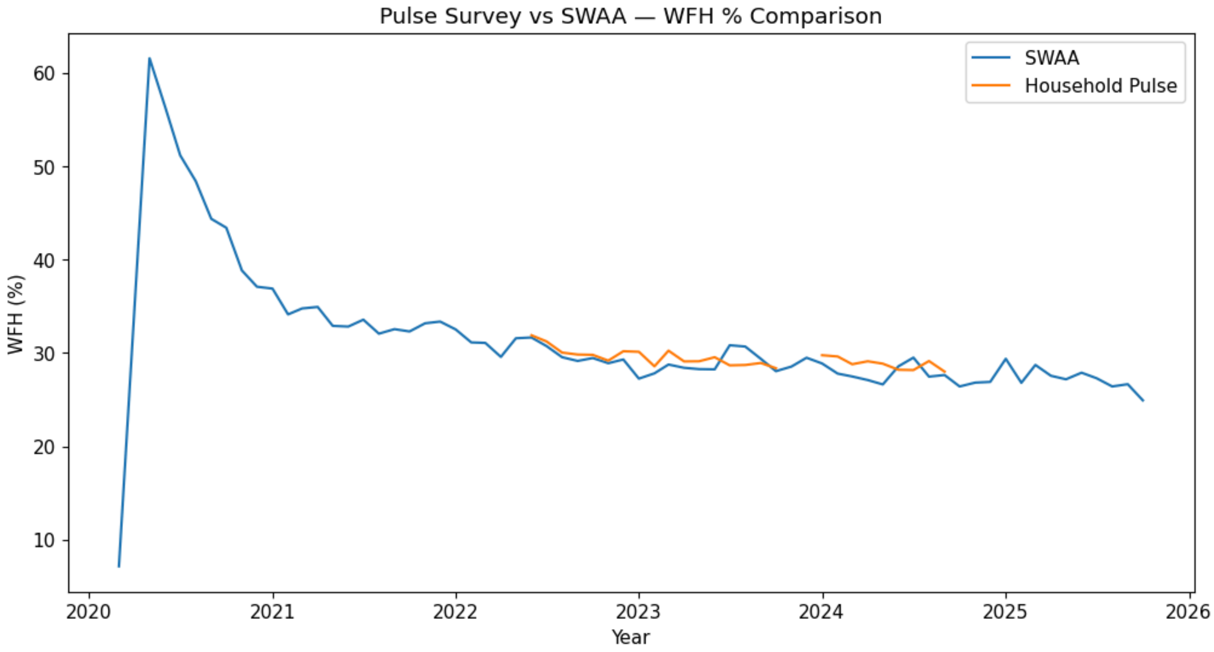
```
df = df_before_during.copy()
df['date'] = pd.to_datetime(df['date'], format='%Y%m')

plt.figure(figsize=(12,6))
plt.plot(df['date'], df['wfhcovid_matquestion'], label='WFH % (SWAA)')
plt.plot(df['date'], df['wfhcovid_frac_HPS'], label='WFH % (Household Pu

plt.title("WFH (% of Full Paid Workdays) Over Time")
plt.xlabel("Year")
plt.ylabel("WFH (% of full paid days)")
plt.legend()
plt.grid()
plt.show()
```



This graph shows how work-from-home levels changed from the start of COVID to today. WFH spiked dramatically in early 2020 as lockdowns forced people home. Over the next few years, the percentage slowly dropped as offices reopened — but it never returned to pre-pandemic levels. Instead, WFH has stabilized around 25–30%, showing that remote work is now a lasting part of how people work.

Pulse Survey vs SWAA — Comparison Line Chart

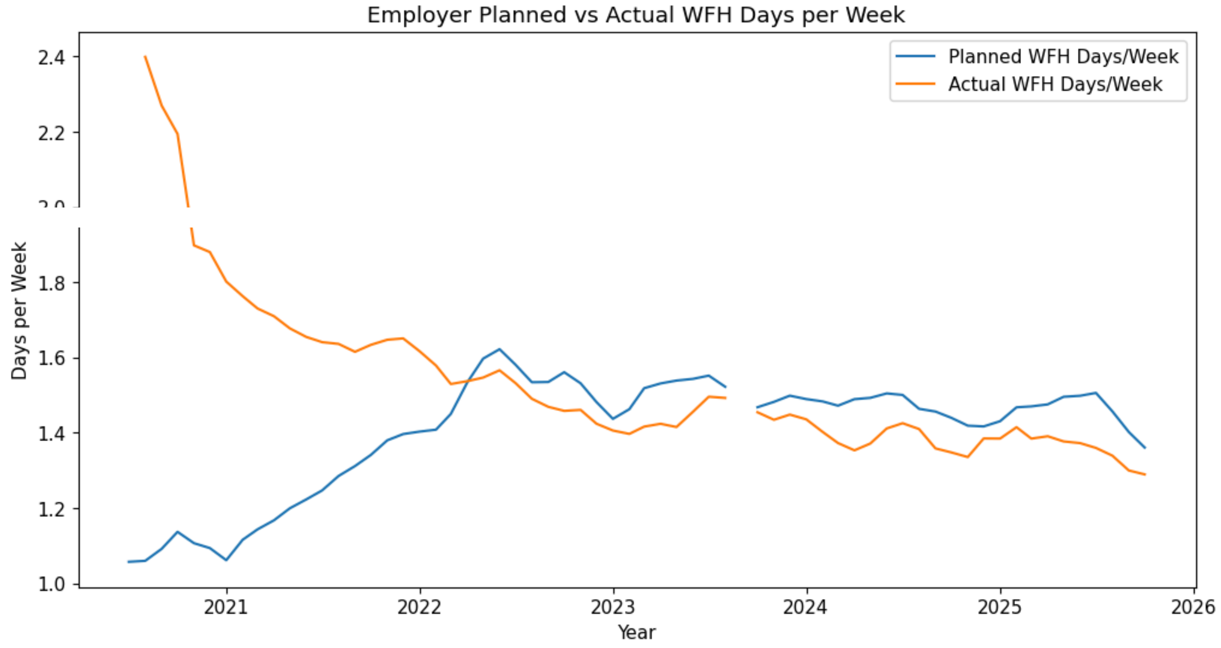
```
plt.figure(figsize=(12,6))
plt.plot(df['date'], df['wfhcovid_matquestion'], label='SWAA')
plt.plot(df['date'], df['wfhcovid_frac_HPS'], label='Household Pulse')

plt.title("Pulse Survey vs SWAA — WFH % Comparison")
plt.xlabel("Year")
plt.ylabel("WFH (%)")
plt.legend()
plt.grid()
plt.show()
```



This comparison shows that two different national surveys tell almost the same story about remote work. Both SWAA and the Household Pulse Survey follow nearly identical trends, which strengthens confidence in the data. The drop after COVID and the stabilization at around 30% is visible in both datasets, showing that remote work has become a permanent and consistent trend across the U.S.

Employer Planned vs Actual WFH Days

```
df_emp = df_emp_plans.copy()
df_emp['date'] = pd.to_datetime(df_emp['date'], format='%Y%m')

plt.figure(figsize=(12,6))
plt.plot(df_emp['date'], df_emp['wfh_days_postCOVID_planMAd'], label='Pl
plt.plot(df_emp['date'], df_emp['wfhcovid_fracmat_hMAd'], label='Actual
plt.title("Employer Planned vs Actual WFH Days per Week")
plt.xlabel("Year")
plt.ylabel("Days per Week")
plt.legend()
plt.grid()
plt.show()
```



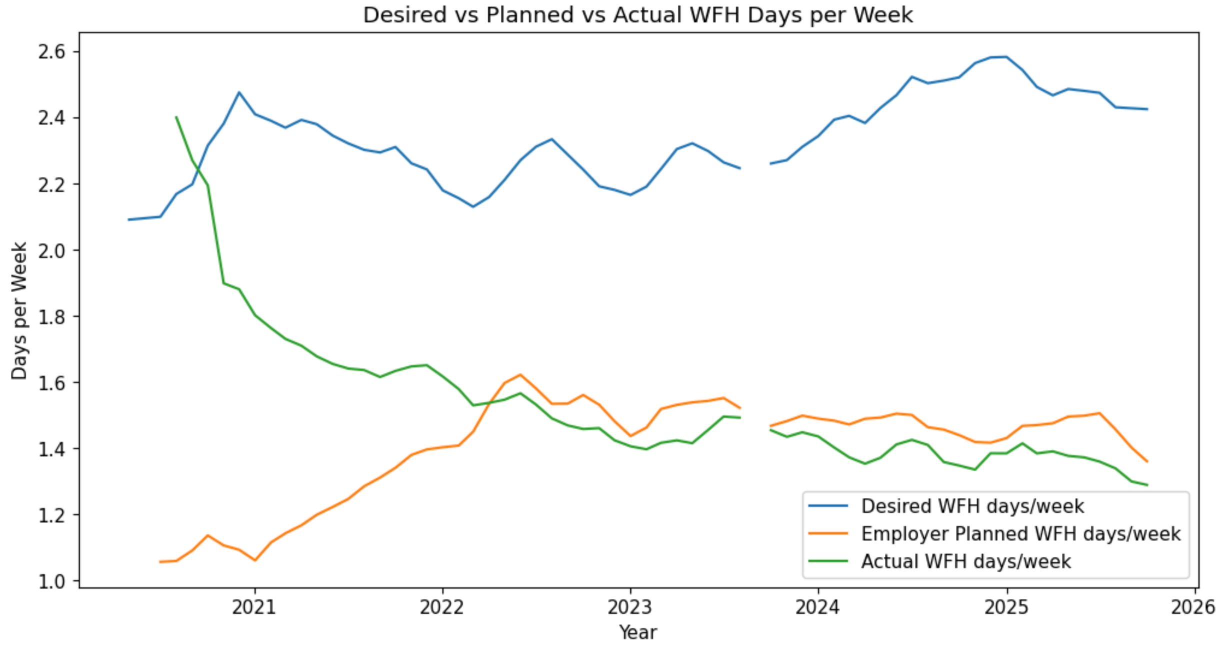
This graph compares how many WFH days companies planned versus how many actually happened. Early on, employees were working from home more often than employers expected. Over time, the two lines move closer, meaning companies gradually adjusted their policies to match real behavior. By 2023 onward, planned and actual days are almost identical — showing that hybrid schedules have stabilized.

Desired vs Planned vs Actual

```
df_worker_desire.columns
```

```
Index(['date', 'wfh_days_postCOVID_desMAd',
'wfh_days_postCOVID_des_eMAd',
       'wfhcovid_fracmat_hMAd', 'wfhcovid_fracmat_eMAd', 'License',
'Citation',
       'Notes'],
      dtype='object')
```

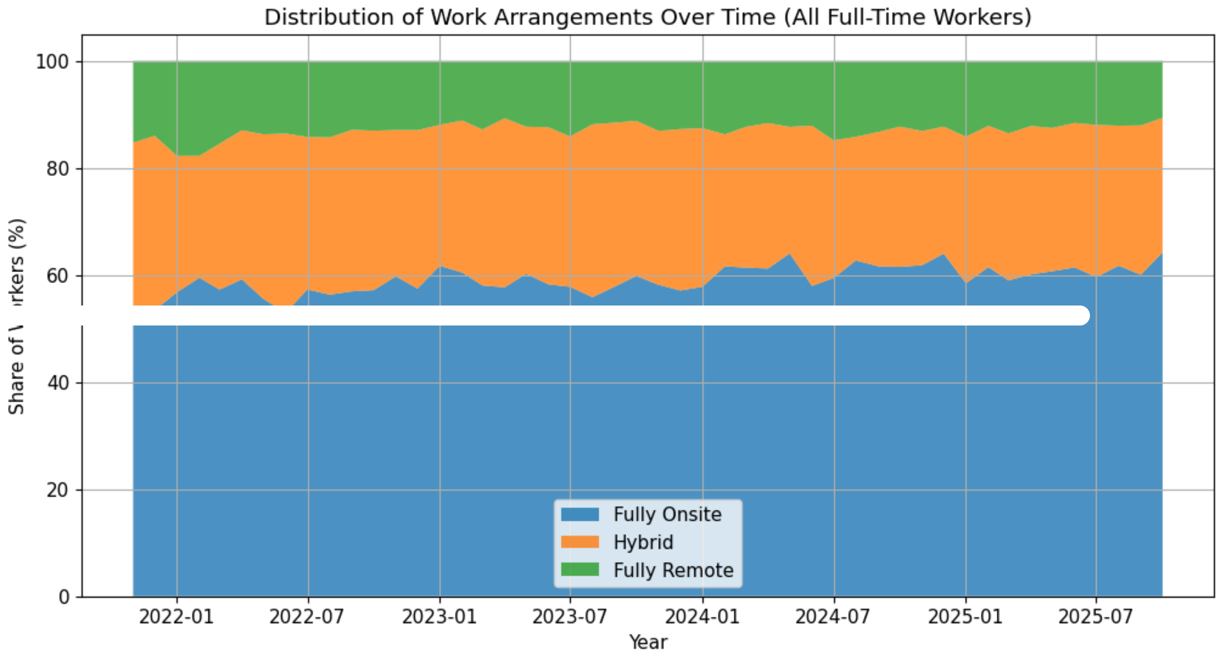
```
df_desire = df_worker_desire.copy()
df_desire['date'] = pd.to_datetime(df_desire['date'], format='%Y%m')

df_plan = df_emp_plans.copy()
df_plan['date'] = pd.to_datetime(df_plan['date'], format='%Y%m')

plt.figure(figsize=(12,6))


plt.plot(df_desire['date'],
         df_desire['wfh_days_postCOVID_desMAd'],
         label='Desired WFH days/week')


plt.plot(df_plan['date'],
         df_plan['wfh_days_postCOVID_planMAd'],
         label='Employer Planned WFH days/week')

plt.plot(df_desire['date'],
         df_desire['wfhcovid_fracmat_hMAd'],
         label='Actual WFH days/week')

plt.title("Desired vs Planned vs Actual WFH Days per Week")
plt.xlabel("Year")
plt.ylabel("Days per Week")
plt.legend()
plt.grid()
plt.show()
```



This graph shows a clear gap between what workers want and what employers offer. Employees consistently prefer more WFH days than companies plan for. Actual WFH tends to fall slightly below employer plans, meaning most people end up working from home less than they would like. Even though the gap narrows slightly, employee preferences for flexibility remain higher than what workplaces currently provide.

Work Arrangement Distribution

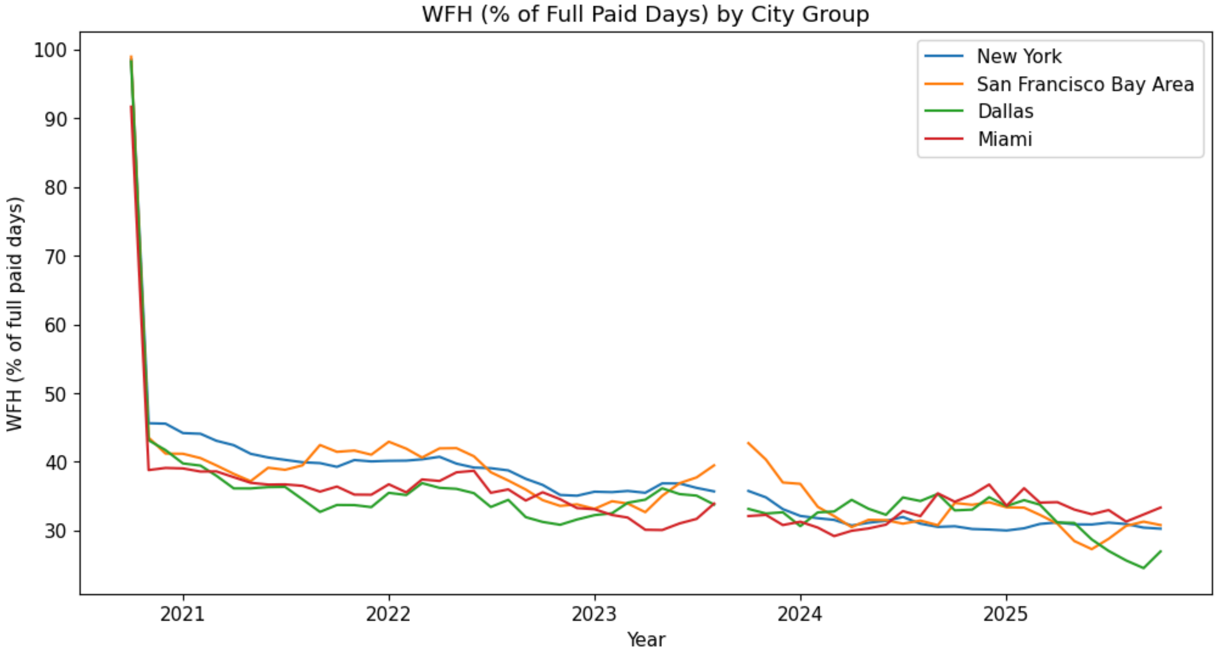
```
df_modes2 = df_modes.copy()
df_modes2['date'] = pd.to_datetime(df_modes2['date'], format='%Y%m')

plt.figure(figsize=(12,6))
plt.stackplot(df_modes2['date'],
              df_modes2['full_onsite_curr'],
              df_modes2['hybrid_curr'],
              df_modes2['full_remote_curr'],
              labels=['Fully Onsite','Hybrid','Fully Remote'],
              alpha=0.8)

plt.title("Distribution of Work Arrangements Over Time (All Full-Time Wo
plt.xlabel("Year")
plt.ylabel("Share of Workers (%)")
plt.legend()
plt.show()
```



This chart shows the overall distribution of work styles — onsite, hybrid, and fully remote — for full-time workers. Most people have returned to fully onsite jobs, but hybrid work has grown into a significant and stable segment, representing about a quarter of all workers. Fully remote work makes up a smaller but still meaningful share. This shows that hybrid work has become a mainstream long-term model.

WFH by City

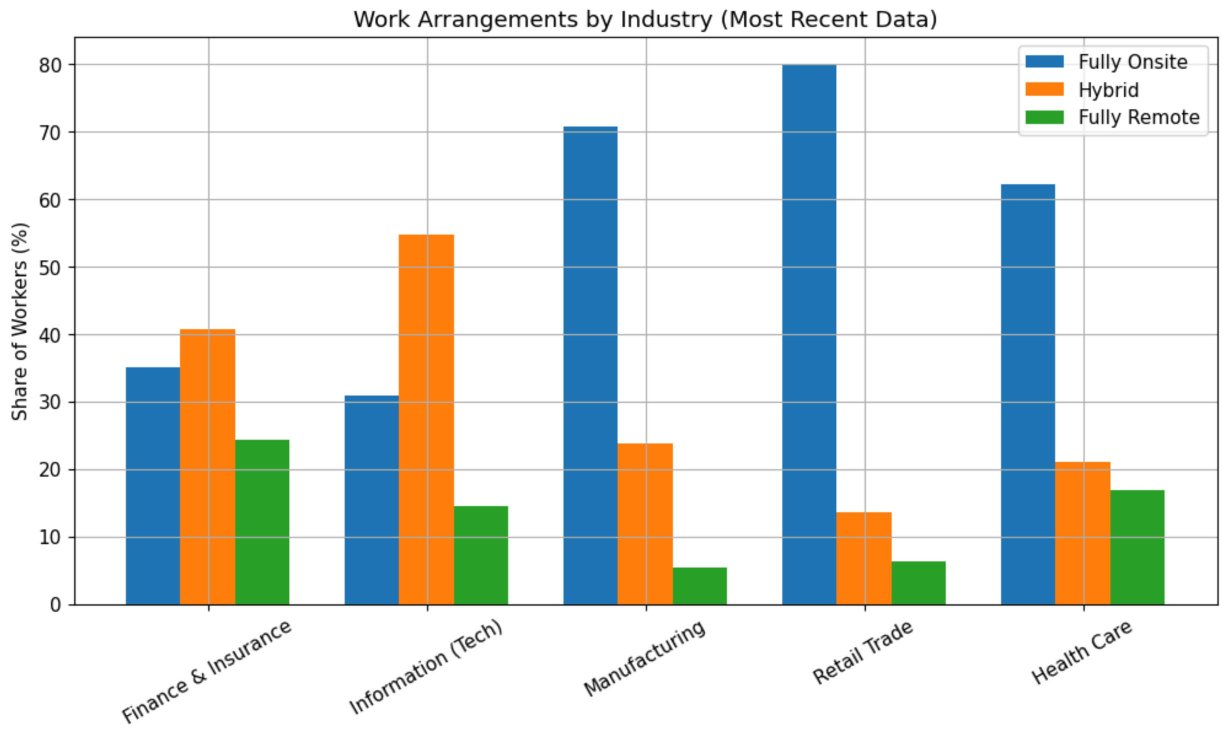
```
df_city2 = df_city.copy()
df_city2['date'] = pd.to_datetime(df_city2['date'], format='%Y%m')

plt.figure(figsize=(12,6))
plt.plot(df_city2['date'], df_city2['wfhcovid_series_MA6_NewYork'], labe
plt.plot(df_city2['date'], df_city2['wfhcovid_series_MA6_BayArea'], labe
plt.plot(df_city2['date'], df_city2['wfhcovid_series_MA6_Dallas'], label
plt.plot(df_city2['date'], df_city2['wfhcovid_series_MA6_Miami'], label=

plt.title("WFH (% of Full Paid Days) by City Group")
plt.xlabel("Year")
plt.ylabel("WFH (% of full paid days)")
plt.legend()
plt.grid()
plt.show()
```



Different cities have very different levels of remote work. Tech-driven, high-cost cities like San Francisco and New York consistently show the highest WFH rates. Mid-sized cities like Dallas and Miami have lower levels but still follow the same overall trend. Over time, all cities decline from the 2020 peak but settle at different levels — highlighting how local industry and commuting patterns shape remote work adoption.

Industry Comparison

```
df_ind = df_industry.copy()

industries = ["Finance & Insurance","Information (Tech)","Manufacturing"

# latest row (most recent data)
last = df_ind.iloc[-1]

onsite = [
    last['full_onsite_finance_insurance'],
    last['full_onsite_information'],
    last['full_onsite_manufacturing'],
    last['full_onsite_retail'],
    last['full_onsite_healthcare']
]

hybrid = [
    last['hybrid_finance_insurance'],
    last['hybrid_information'],
    last['hybrid_manufacturing'],
    last['hybrid_retail'],
    last['hybrid_healthcare']
]

remote = [
    last['full_remote_finance_insurance'],
    last['full_remote_information'],
    last['full_remote_manufacturing'],
    last['full_remote_retail'],
    last['full_remote_healthcare']
]

x = np.arange(len(industries))
width = 0.25

plt.figure(figsize=(12,6))
plt.bar(x - width, onsite, width, label='Fully Onsite')
plt.bar(x, hybrid, width, label='Hybrid')
plt.bar(x + width, remote, width, label='Fully Remote')

plt.xticks(x, industries, rotation=30)
plt.title("Work Arrangements by Industry (Most Recent Data)")
plt.ylabel("Share of Workers (%)")
plt.legend()
plt.show()
```

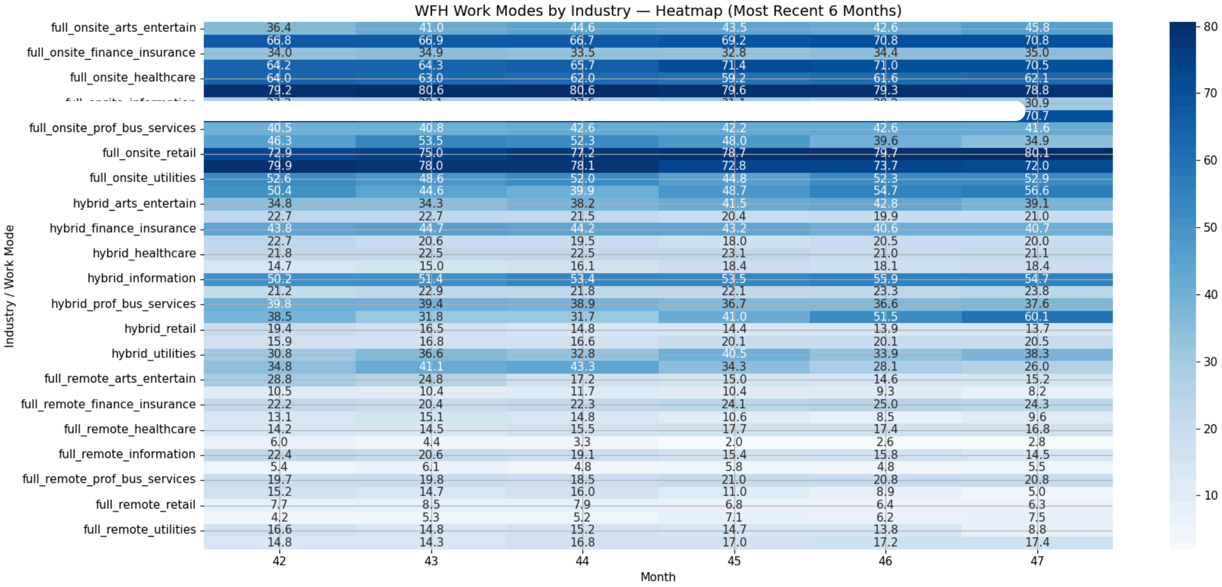


This chart compares how work arrangements vary across industries. Knowledge-based sectors like tech and finance have the highest share of hybrid and remote workers. Meanwhile, manufacturing, retail, and healthcare remain mostly onsite because their work requires physical presence. The chart clearly shows that remote work opportunities depend heavily on the type of job and industry.

Industry Heatmap

```
import seaborn as sns
import matplotlib.pyplot as plt
```

```
industry_cols = [col for col in df_industry.columns
                 if ('full_onsite' in col
                     or 'hybrid' in col
                     or 'full_remote' in col)]

df_heat = df_industry[industry_cols].dropna(axis=1, how="all")

heatmap_df = df_heat.tail(6)

plt.figure(figsize=(18, 8))
sns.heatmap(heatmap_df.T, annot=True, cmap="Blues", fmt=".1f")
plt.title("WFH Work Modes by Industry — Heatmap (Most Recent 6 Months)",
plt.xlabel("Month")
plt.ylabel("Industry / Work Mode")
plt.tight_layout()
plt.show()
```



The heatmap shows how much different industries rely on remote work. Darker shades indicate higher WFH percentages. Tech and finance typically show the highest values, while manufacturing and retail remain low. This visual highlights that remote work isn’t evenly distributed — it strongly favors industries where tasks can be done digitally.

Work-From-Home vs Back-to-Office: A Data-Driven Analysis

Introduction

Remote and hybrid work have reshaped how organizations think about productivity, flexibility, and well-being. Using data from the WFH Research project, this notebook explores long-term trends in WFH adoption, worker preferences, employer planning, and