Public health awareness campaign analysis using IBM Cognos

Building a public health awareness campaign analysis using IBM Cognos for visualization involves several key steps, including defining analysis objectives, collecting campaign data, and processing and cleaning the data. Here's a step-by-step guide on how to get started:

1. Define Analysis Objectives:

Start by clearly defining the objectives of your public health awareness campaign analysis. What specific insights are you trying to gain from the data? For example:
  - Assess the effectiveness of the campaign in terms of reach and engagement.
  - Identify target demographics that responded well to the campaign.
  - Evaluate the impact of the campaign on public health awareness and behaviour.
  - Measure the return on investment (ROI) of the campaign.

2.Implementing a dataset

Kaggle dataset link:https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey

	Timestamp	Age	Gender	Country	state	self_employed	family_history	treatment	work_interfere	no_employees	...	leave
0	2014-08-27 11:29:31	37	Female	United States	IL	NaN	No	Yes	Often	6-25	...	Some easy
1	2014-08-27 11:29:37	44	M	United States	IN	NaN	No	No	Rarely	More than 1000	...	Don't know
2	2014-08-27 11:29:44	32	Male	Canada	NaN	NaN	No	No	Rarely	6-25	...	Some diffic
3	2014-08-27 11:29:46	31	Male	United Kingdom	NaN	NaN	Yes	Yes	Often	26-100	...	Some diffic
4	2014-08-27 11:30:22	31	Male	United States	TX	NaN	No	No	Never	100-500	...	Don't know

3. Collect Campaign Data:

To collect data for your analysis, you'll need access to the campaign data source. This may involve reaching out to relevant public health organisations, government agencies, or partners who have run the campaign. Ensure you have permission to use and analyse the data. The data you collect may include:

- Social media engagement metrics (likes, shares, comments).
- Website traffic and click-through rates.
- Surveys or questionnaires to assess public awareness and behaviour change.
- Demographic data of the campaign's audience.

```python
In [1]:   #imports necessary libraries to do basic things on the dataset
          import pandas as pd
          import numpy as np

          import seaborn as sns
          import matplotlib.pyplot as plt



          print('Successfully imported')

          Successfully imported
```

3. Process and Clean Data:

Data processing and cleaning are critical to ensuring the quality and accuracy of your analysis. Here are some steps to follow:

a. Data Integration: If your campaign data is stored in different formats or sources, integrate them into a single dataset. This may involve using ETL (Extract, Transform, Load) tools.

b. Data Cleaning:
- Handle missing data: Identify and deal with missing values, either by imputation or removal.
- Remove duplicates: Eliminate duplicate records.
- Data format standardisation: Ensure that date formats, units of measurement, and naming conventions are consistent.

```
#Check the dataset for missing data
if data.isnull().sum().sum() == 0 :
    print ('There is no missing data in our dataset')
else:
    print('There is {} missing data in our dataset '.format(data.isnull().sum().sum()))
```

There is 1892 missing data in our dataset

```
#Check our missing data from which columns and how many unique features they have.
frame = pd.concat([data.isnull().sum(), data.nunique(), data.dtypes], axis = 1, sort= False)
frame
```

c. Data Transformation:
    - Convert data types: Ensure that data types are appropriate for analysis (e.g., dates as date objects, numbers as numeric types).
    - Create calculated fields: Generate new variables if needed (e.g., calculate engagement rates).
    - Aggregation: Summarise data as needed, e.g., daily, weekly, or monthly aggregates.

|	0	1	2
Timestamp	0	1246	object
Age	0	53	int64
Gender	0	49	object
Country	0	48	object
state	515	45	object
self_employed	18	2	object
family_history	0	2	object
treatment	0	2	object
work_interfere	264	4	object
no_employees	0	6	object
remote_work	0	2	object
tech_company	0	2	object
benefits	0	3	object
care_options	0	3	object
wellness_program	0	3	object
seek_help	0	3	object
anonymity	0	3	object
leave	0	5	object
mental_health_consequence	0	3	object
phys_health_consequence	0	3	object
coworkers	0	3	object
supervisor	0	3	object
mental_health_interview	0	3	object
phys_health_interview	0	3	object
mental_vs_physical	0	3	object
obs_consequence	0	2	object
comments	1095	160	object

d. Data Quality Check:
   - **Check for outliers and anomalies that may affect analysis.**
   - **Validate data against the defined analysis objectives.**

4. Create Visualizations and Reports:

   With your data in IBM Cognos, you can start building visualizations and reports to address your analysis objectives. You can use various chart types, tables, and graphs to present the data effectively.

```
In [6]:
#Plot **work_interfere**
ax = sns.countplot(data = data , x = 'work_interfere');
#Add the value of each parametr on the Plot
ax.bar_label(ax.containers[0]);
```
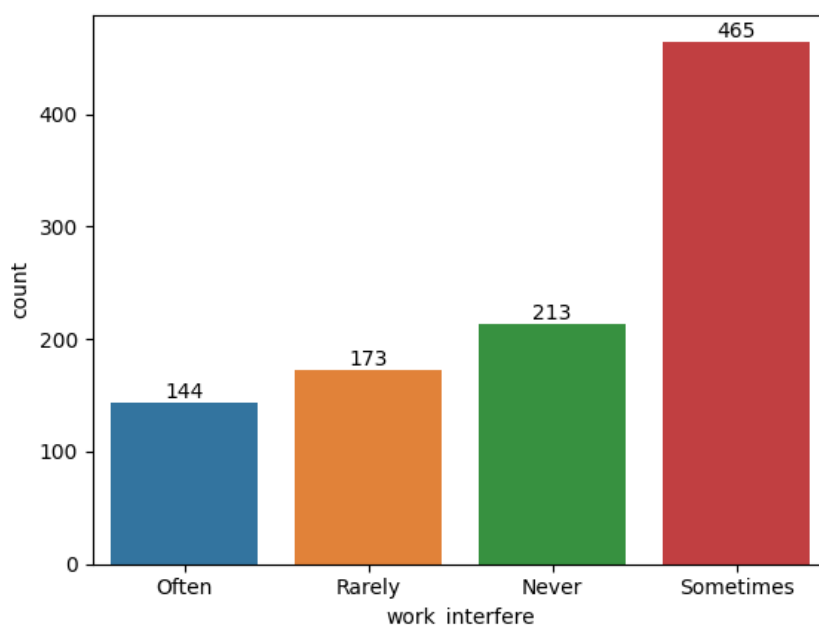

5. Analyze the Data:

   Analyze the data to draw insights and conclusions related to your campaign's effectiveness. Use features within IBM Cognos to perform statistical analysis or apply business intelligence techniques to discover patterns and trends.

```python
from sklearn.impute import SimpleImputer
import numpy as np
columns_to_drop = ['state', 'comments', 'Timestamp']
for column in columns_to_drop:
    if column in data.columns:
        data = data.drop(columns=[column])


# Fill in missing values in work_interfere column
data['work_interfere'] = np.ravel(SimpleImputer(strategy = 'most_frequent').fit_transform(data['work
_interfere'].values.reshape(-1,1)))
data['self_employed'] = np.ravel(SimpleImputer(strategy = 'most_frequent').fit_transform(data['self_
employed'].values.reshape(-1,1)))

data.head()
```

	Age	Gender	Country	self_employed	family_history	treatment	work_interfere	no_employees	remote_work	tech_company
0	37	Female	United States	No	No	Yes	Often	6-25	No	Yes
1	44	M	United States	No	No	No	Rarely	More than 1000	No	No
2	32	Male	Canada	No	No	No	Rarely	6-25	No	Yes
3	31	Male	United Kingdom	No	Yes	Yes	Often	26-100	No	Yes
4	31	Male	United States	No	No	No	Never	100-500	Yes	Yes

6. Share Insights:

  Finally, share the insights and findings from your analysis with relevant stakeholders, including campaign organizers, public health officials, or other decision-makers. Use the reports and visualizations created in IBM Cognos for this purpose.

```
ax = sns.countplot(data=data, x='work_interfere');
ax.bar_label(ax.containers[0]);
```
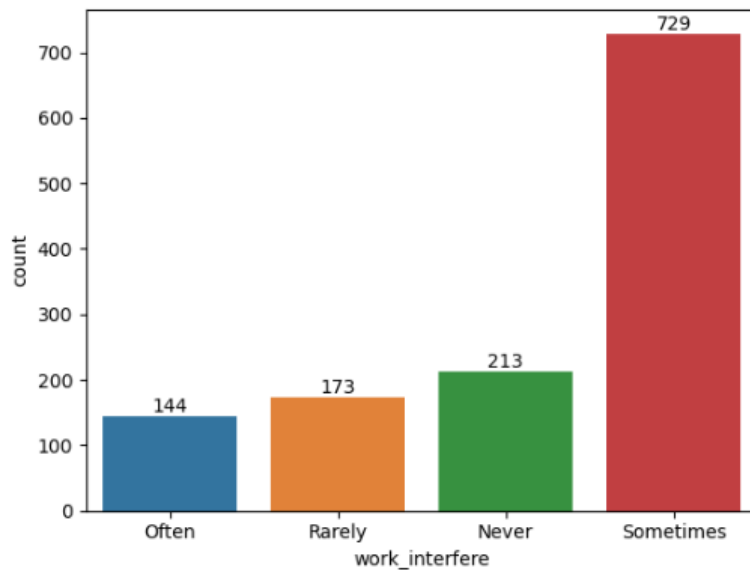

7. Iterate and Refine:

  Data analysis is an iterative process. If your analysis reveals areas for improvement in the campaign, work with stakeholders to refine strategies and potentially run new campaigns.

Remember to document your analysis process, including the data sources, data cleaning steps, and the rationale behind your analysis choices. This documentation will be valuable for future reference and for ensuring the transparency and reproducibility of your work.