

Clinical Trial Data Analysis Project

Exploratory Data Analysis using Python

Prepared by: Ritika Shah

Python Code:

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

Python Code:

```
## Clinical Trial Trends & Pharma Analytics ##This project analyzes real-world clinical trial data to understand ##research trends across trial phases, sponsors, diseases, and outcomes.
```

Python Code:

```
df=pd.read_csv("clinical.csv.csv")
```

Python Code:

```
df.head()
```

Python Code:

```
df.info()
```

Python Code:

```
df.describe()
```

Python Code:

```
df.isnull().sum()
```

Python Code:

```
df['Enrollment'] = df['Enrollment'].fillna(0)
```

Python Code:

```
df=df.drop_duplicates( subset='NCT' )
```

Python Code:

```
df.columns=df.columns.str.strip().str.lower().str.replace(" ","-")
```

Python Code:

```
df['phase'].value_counts()
```

Python Code:

```
df['status'].value_counts()
```

Python Code:

```
#VISUALIZE THE PHASE DISTRIBUTION plt.figure(figsize=(8,5))  
sns.countplot(x='phase',data=df,order=df['phase'].value_counts().index)  
plt.title('DISTRIBUTION OF CLINICAL TRIAL PHASE') plt.xlabel("Phase") plt.ylabel("Number of Trial") plt.xticks(rotation=45) plt.show()
```

Python Code:

```
#VISUALIZATION OF STATUS plt.figure(figsize=(8,5))
sns.countplot(x='status',data=df,order=df['status'].value_counts().index)
plt.title('DISTRIBUTION OF CLINICAL TRAITS STATUS') plt.xlabel('Status') plt.ylabel('Number of Trials') plt.xticks(rotation=45) plt.show()
```

Python Code:

```
#cross tab between status and phase phase_status= pd.crosstab(df['phase'],df['status'])
print(phase_status)
```

Python Code:

```
#VISUALIZE THE PHASE AND STATUS BY HEATMAP plt.figure(figsize=(10,6))
sns.heatmap(phase_status, annot=True, fmt='d', cmap='YlGnBu') plt.title('Phase vs Status of Clinical Trials') plt.xlabel('Status') plt.ylabel('Phase') plt.show()
```

Python Code:

```
df['enrollment'].describe()
```

Python Code:

```
df_enroll = df[df['enrollment'].notna()&(df['enrollment']>0)]
```

Python Code:

```
avg_enrollment_phase=df_enroll.groupby('phase')['enrollment'].mean().sort_values(ascending=False) print("Average Enrollment by Phase:") print(avg_enrollment_phase)
```

Python Code:

```
plt.figure(figsize=(8,5))
sns.barplot(x=avg_enrollment_phase.index,y=avg_enrollment_phase.values) plt.title('Average Enrollment by Clinical Trial Phase') plt.xlabel('Phase') plt.ylabel('Average Number of Participants') plt.xticks(rotation =45) plt.show()
```

Python Code:

```
avg_enrollment_status=df_enroll.groupby('status')['enrollment'].mean().sort_values(ascending=False) print('Average Enrollment by Status') print(avg_enrollment_status)
```

Python Code:

```
plt.figure(figsize=8,5)
sns.barplot(x=avg_enrollment_status.index,y=avg_enrollment_status.values)
plt.title('Average Enrollment by Trial Status') plt.xlabel('Status') plt.ylabel('Average Number by Participants') plt.xticks(rotation=45) plt.show()
```

Python Code:

```
enrollment_phase_status=df_enroll.pivot_table(index='phase',columns='status',values='enrollment',aggfunc='mean') plt.figure(figsize=(10,7)) sns.heatmap(enrollment_phase_status, annot=True, fmt=" .1f",cmap='YlOrRd') plt.title('Average Enrollment by Phase and Status') plt.xlabel('Status') plt.ylabel('Phase') plt.show()
```

Python Code:

```
#Enrollment Comparrison plt.figure(figsize=(8,6))
sns.scatterplot(x='phase',y='enrollment',data=df) plt.title('Enrollment Distribution BY
Phase') plt.xlabel('Clinical Trial Phase') plt.ylabel('Number Of Participant') plt.show()
```

Python Code:

```
#Enrollment Comparrison BY Status plt.figure(figsize=(9,16))
sns.scatterplot(x='status',y='enrollment',data=df) plt.title('Enrollment Distribution By
Status') plt.xlabel('Trial Status') plt.ylabel('Number Of Participant')
plt.xticks(rotation=45) plt.show()
```