

Project Management Analysis

Using the functions and capabilities of Python and the data analysis libraries, we tackle the analysis of the Project Management dataset. We will explore the key attributes: Project Name, Project Description, Project Type, Project Manager, Region, Department, Project Cost, Project Benefit, Complexity, Status, Completion, Phase, Year, Month, Start Date & End Date to gain insight on what influences the projects' progress and state. Visualizing the findings using data visualization techniques and conducting analysis to identify: trends, patterns and correlations within the dataset to provide a satisfactory report.

Import Library

+ Code + Markdown

```
1] import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

Python

Import CSV File

```
2] df = pd.read_csv("12_Project Management Analysis.csv")
```

Python

Data Preprocessing

`.head()`

Using `.head()` to display the first 5 rows of our dataset.

```
df.head()
```

	Project Name	Project Description	Project Type	Project Manager	Region	Department	Project Cost	Project Benefit	Complexity	Status	Completion%	Phase	Year	Month	Start Date	End Date
0	Rhinestone	Associations Now Is A Casual Game To Teach You...	INCOME GENERATION	Yael Wilcox	North	Admin & BI	3648615	8443980	High	In - Progress	77%	Phase 4 - Implement	2021	2	2/1/2021	6/1/2021
1	A Triumph Of Softwares	Is A Fully Managed Content Marketing Software ...	INCOME GENERATION	Brenda Chandler	West	eCommerce	4018835	9012225	High	Cancelled	80%	Phase 2 - Develop	2021	3	3/1/2021	6/1/2021
2	The Blue Bird	Most Content Marketers Know The Golden Rule: Y...	INCOME GENERATION	Nyasia Hunter	North	Warehouse	4285483	9078339	High	Completed	100%	Phase 4 - Implement	2021	3	3/1/2021	6/1/2021
3	Remembering Our Ancestors	Utilize And Utilizes (Verb Form) The Open, Inc...	PROCESS IMPROVEMENT	Brenda Chandler	East	Sales and Marketing	5285864	8719006	High	Cancelled	75%	Phase 5 - Measure	2021	3	3/1/2021	6/1/2021
4	Skyhawks	Is A Solution For Founders Who Want To Win At ...	WORKING CAPITAL IMPROVEMENT	Jaylyn Mckenzie	East	eCommerce	5785601	8630148	High	Completed	100%	Phase 1 - Explore	2021	3	3/1/2021	6/1/2021

`.tail()`

Using `.tail()` to show the last 5 rows of the dataset.

Using .tail() to show the last 5 rows of the dataset.

```
df.tail()
```

Python

	Project Name	Project Description	Project Type	Project Manager	Region	Department	Project Cost	Project Benefit	Complexity	Status	Completion%	Phase	Year	Month	Start Date	End Date
94	Strive Training	Was Built To Help Founders Create Optimized Co...	WORKING CAPITAL IMPROVEMENT	Nyasia Hunter	South	Supply Chain	5259436	8817917	Medium	On - Hold	80%	Phase 2 - Develop	2025	8	8/1/2025	11/1/2025
95	Debug Entity	In This Ecosystem, Association Content Is Simp...	INCOME GENERATION	Kamari Norris	North	Warehouse	4790417	8872443	Medium	In - Progress	73%	Phase 4 - Implement	2025	9	9/1/2025	12/1/2025
96	Made By Me	With 15 Five, We Take The Guesswork Out Of Con...	PROCESS IMPROVEMENT	Yael Wilcox	West	Supply Chain	4283481	8895152	Low	Completed	100%	Phase 3 - Plan	2025	11	11/1/2025	3/1/2026
97	Revolution	Was Founded To Help Founders And Entrepreneurs...	COST REDUCTION	Jaylyn Mckenzie	East	eCommerce	4606575	8658343	High	In - Progress	77%	Phase 4 - Implement	2025	11	11/1/2025	3/1/2026
98	7Th Annual Workshop	Welcome To The Future Of Content Creation. The...	WORKING CAPITAL IMPROVEMENT	Nyasia Hunter	West	Sales and Marketing	5054482	8422578	High	In - Progress	83%	Phase 3 - Plan	2025	12	12/1/2025	3/1/2026

.shape

With .shape, we can get the total rows and columns of the dataset.

```
df.shape
```

Python

(99, 16)

.columns

.columns allow us to identify all columns present in the dataset.

```
df.columns
```

Python

```
Index(['Project Name', 'Project Description', 'Project Type',  
      'Project Manager', 'Region', 'Department', 'Project Cost',  
      'Project Benefit', 'Complexity', 'Status', 'Completion%', 'Phase',  
      'Year', 'Month', 'Start Date', 'End Date'],  
      dtype='object')
```

.dtypes

With .dtypes, we can identify the data types assigned to each column

```
df.dtypes
```

Python

```
Project Name      object  
Project Description  object  
Project Type      object  
Project Manager    object  
Region            object  
Department         object  
Project Cost       int64  
Project Benefit     int64  
Complexity         object
```

.unique()

.unique() shows the unique values in a specified column.

```
df['Project Type'].unique()  
  
... array(['INCOME GENERATION', 'PROCESS IMPROVEMENT',  
        'WORKING CAPITAL IMPROVEMENT', 'COST REDUCTION'], dtype=object)
```

Python

.nunique()

.nunique() on the other hand provides us the number of unique values in each columns.

```
df.nunique()  
  
... Project Name      99  
Project Description  95  
Project Type         4  
Project Manager      7  
Region              4  
Department          5  
Project Cost        99  
Project Benefit      99  
Complexity           3  
Status              4  
Completion%         22  
Phase               5  
Year                5  
Month               12  
Start Date          49  
End Date            43  
dtype: int64
```

Python

.describe()

Shows the count, mean, median, etc. of columns with Int64 datatypes.

```
df.describe()  
  
...  
   Project Cost  Project Benefit  Year  Month  
count  9.900000e+01  9.900000e+01  99.000000  99.000000  
mean    4.156649e+06  8.828178e+06  2022.747475  7.151515  
std     1.076544e+06  2.164019e+05   1.402210  3.211471  
min     2.418301e+06  8.422578e+06  2021.000000  1.000000  
25%     3.251948e+06  8.656248e+06  2022.000000  4.500000  
50%     4.172827e+06  8.846243e+06  2022.000000  7.000000  
75%     5.063288e+06  9.019234e+06  2024.000000  10.000000  
max     5.974815e+06  9.165877e+06  2025.000000  12.000000
```

Python

.value_counts()

Returns the number of all unique values in a column.

```
df['Project Type'].value_counts()  
  
Python
```

```
.. Project Type
INCOME GENERATION      27
PROCESS IMPROVEMENT    25
WORKING CAPITAL IMPROVEMENT 25
COST REDUCTION         22
Name: count, dtype: int64
```

`.isnull()`

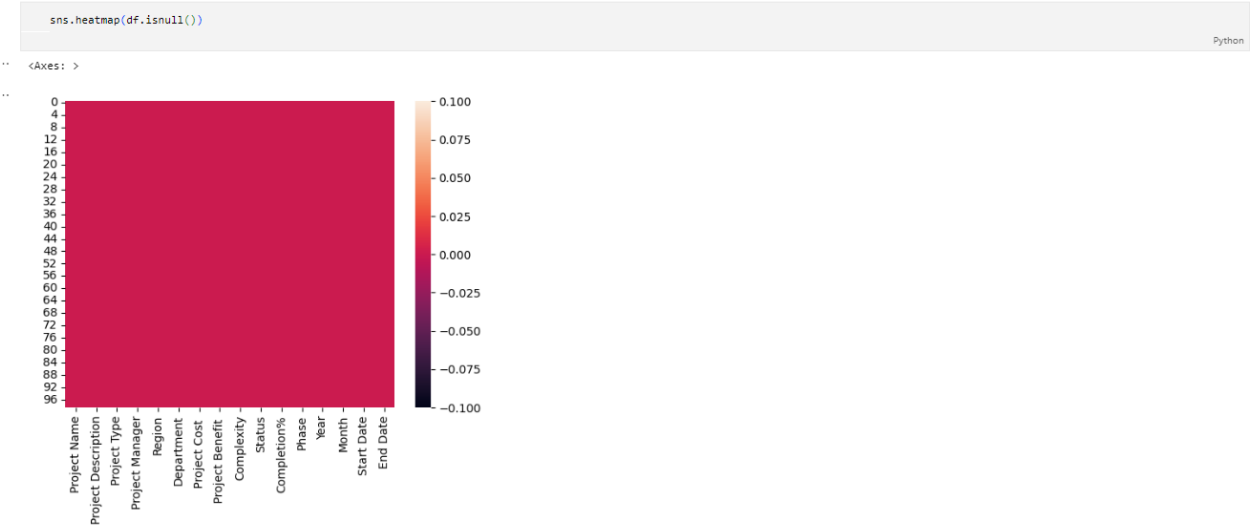
Checks for null values.

df.isnull()

Python

	Project Name	Project Description	Project Type	Project Manager	Region	Department	Project Cost	Project Benefit	Complexity	Status	Completion%	Phase	Year	Month	Start Date	End Date
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
...
94	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
95	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
96	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
97	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
98	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False

99 rows x 16 columns



Data Analysis

```
print(df)
```

	Project Name \				
0	Rhinestone				
1	A Triumph Of Softwares				
2	The Blue Bird				
3	Remembering Our Ancestors				
4	Skyhawks				
..	...				
94	Strive Training				
95	Debug Entity				
96	Made By Me				
97	Revolution				
98	7Th Annual Workshop				

	Project Description \				
0	Associations Now Is A Casual Game To Teach You...				
1	Is A Fully Managed Content Marketing Software ...				
2	Most Content Marketers Know The Golden Rule: Y...				
3	Utilize And Utilizes (Verb Form) The Open, Inc...				
4	Is A Solution For Founders Who Want To Win At ...				
..	...				
94	Was Built To Help Founders Create Optimized Co...				
95	In This Ecosystem, Association Content Is Simp...				
96	With 15 Five, We Take The Guesswork Out Of Con...				
97	Was Founded To Help Founders And Entrepreneurs...				
98	Welcome To The Future Of Content Creation. The...				
...					
97	Phase 4 - Implement	2025	11	11/1/2025	3/1/2026
98	Phase 3 - Plan	2025	12	12/1/2025	3/1/2026

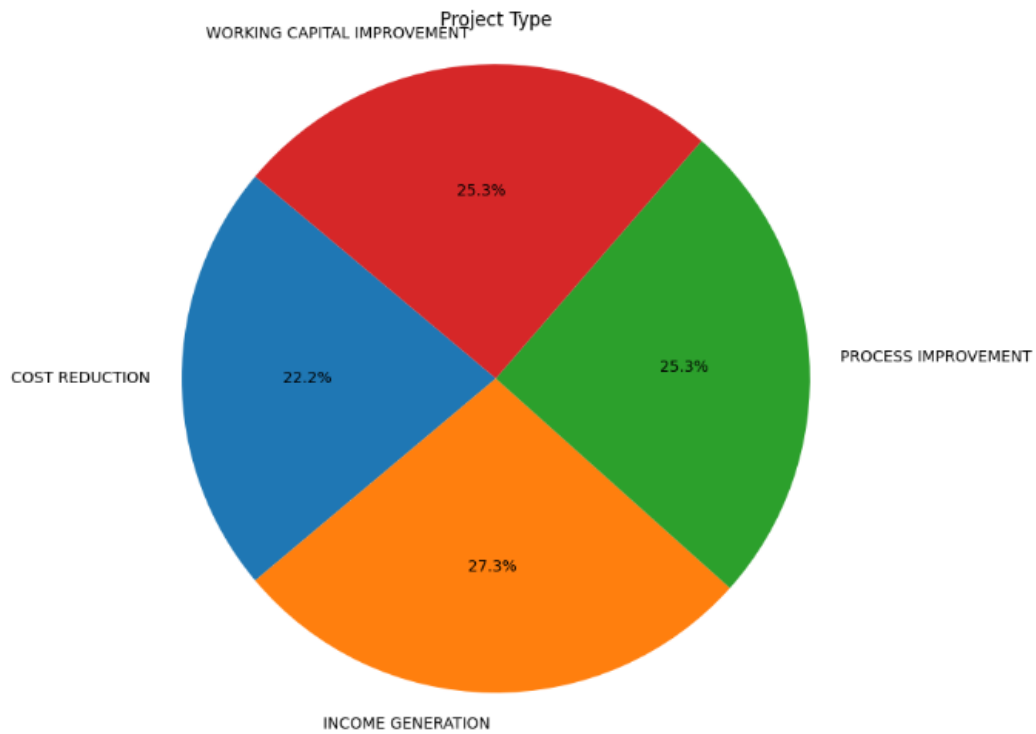
[99 rows x 16 columns]

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings](#).

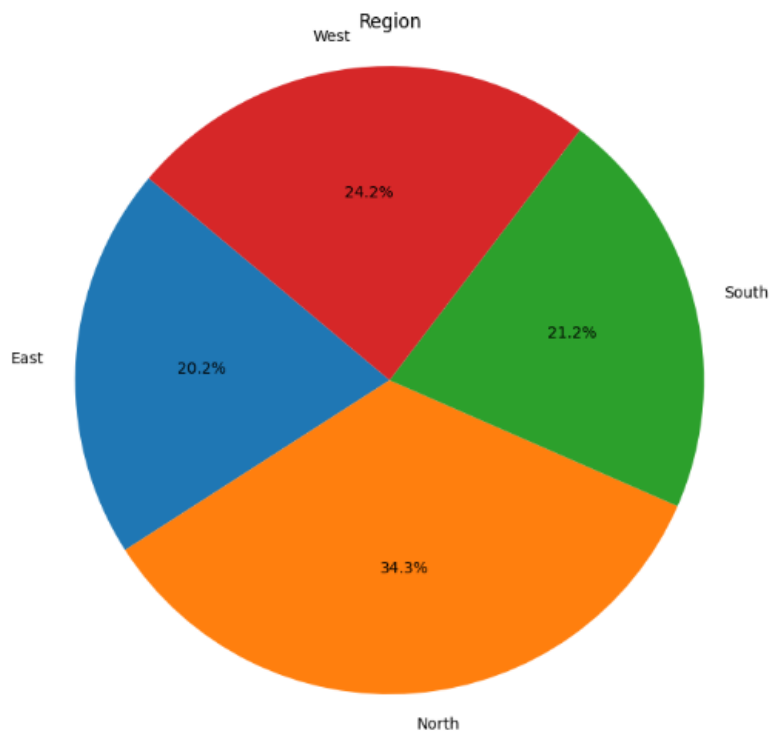
Data Visualization

```
ptype = df.groupby('Project Type').size()
reg = df.groupby('Region').size()
dep = df.groupby('Department').size()
stats = df.groupby('Status').size()
comp = df.groupby('Complexity').size()
pha = df.groupby('Phase').size()

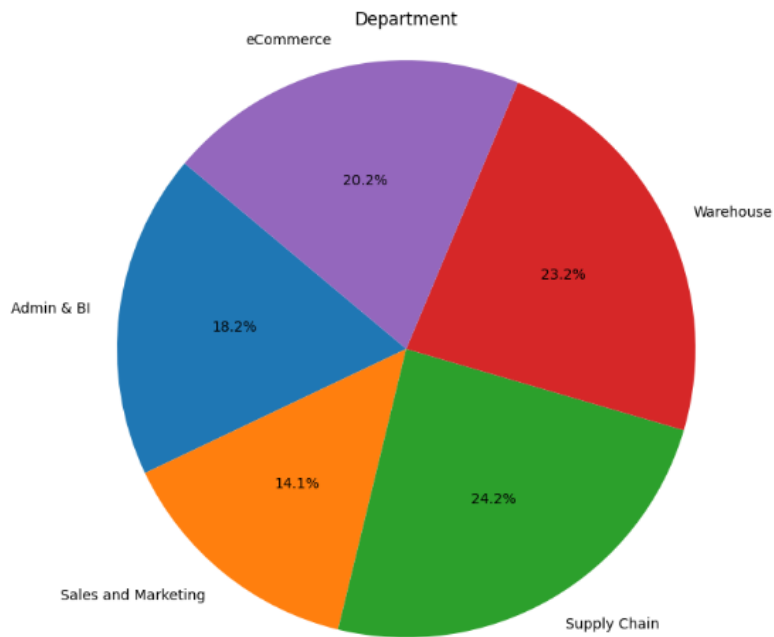
plt.figure(figsize=(8,8))
plt.pie(ptype, labels=ptype.index, autopct='%1.1f%%', startangle=140)
plt.title('Project Type')
plt.axis('equal')
plt.show()
```



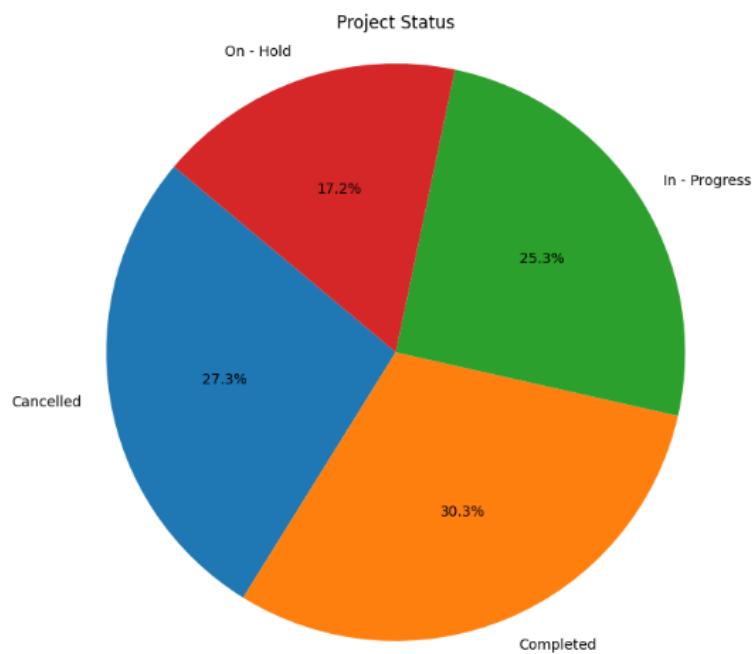
```
plt.figure(figsize=(8,8))
plt.pie(reg, labels=reg.index, autopct='%1.1f%%', startangle=140)
plt.title('Region')
plt.axis('equal')
plt.show()
```



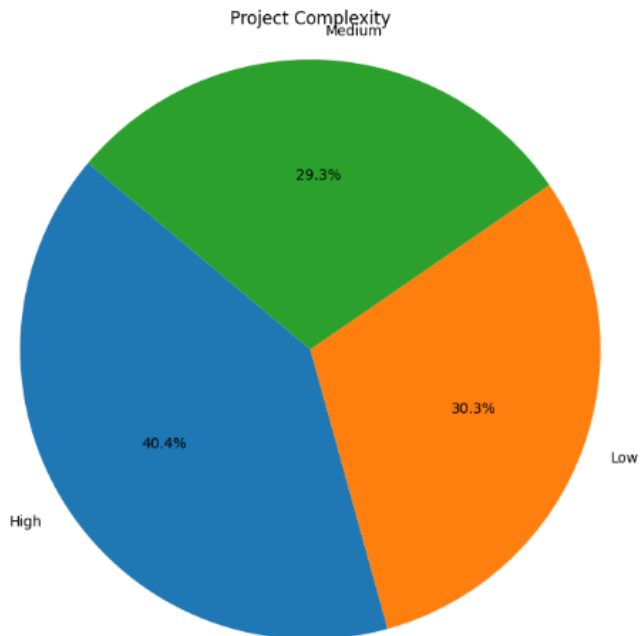
```
plt.figure(figsize=(8,8))
plt.pie(dep, labels=dep.index,autopct='%1.1f%%', startangle=140)
plt.title('Department')
plt.axis('equal')
plt.show()
```



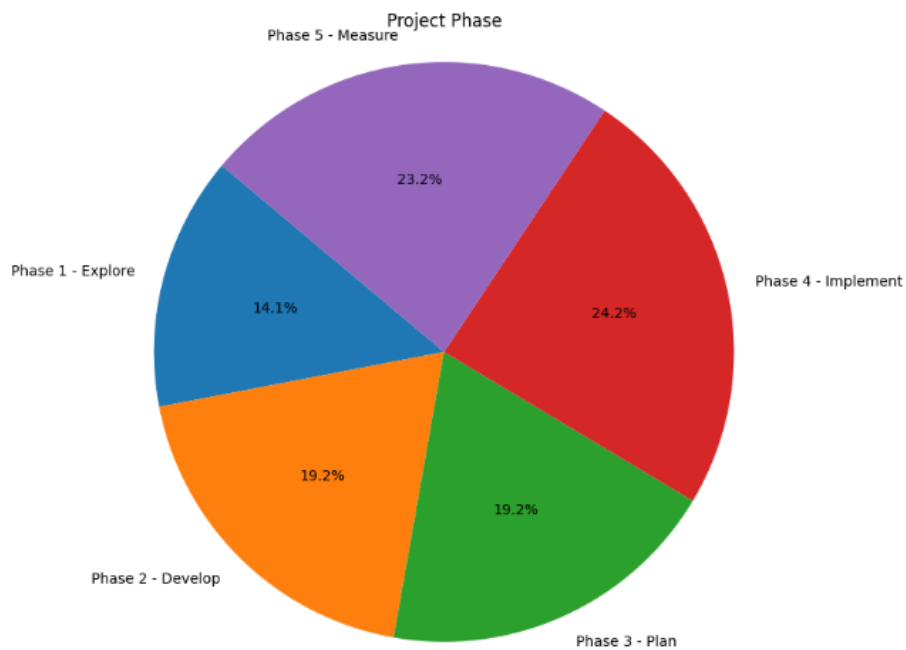
```
plt.figure(figsize=(8,8))
plt.pie(stats, labels=stats.index,autopct='%1.1f%%', startangle=140)
plt.title('Project Status')
plt.axis('equal')
plt.show()
```



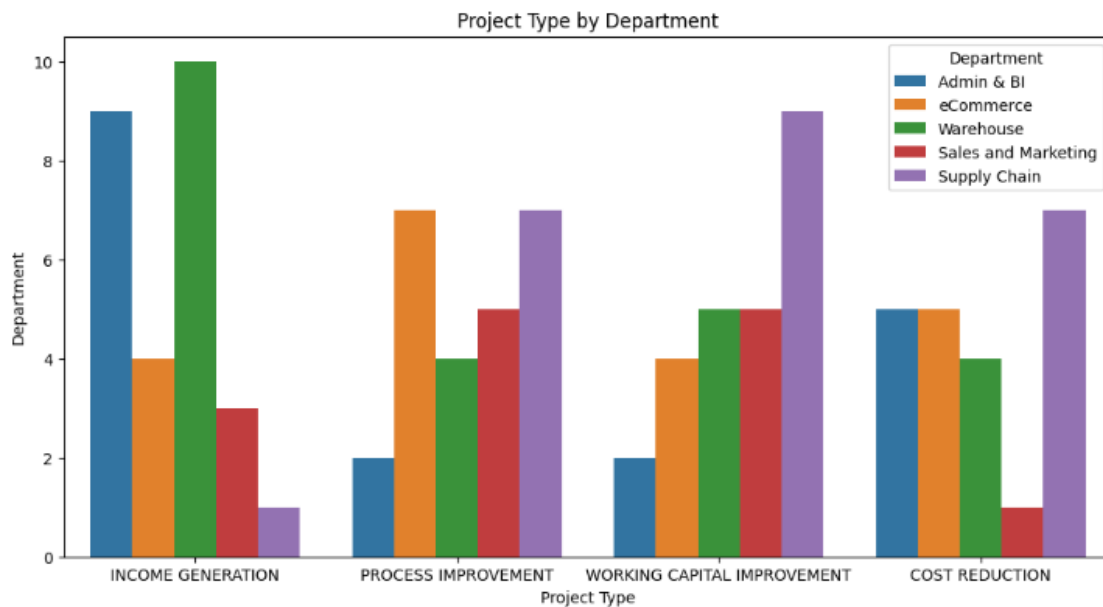
```
> plt.figure(figsize=(8,8))
plt.pie(comp, labels=comp.index,autopct='%1.1f%%', startangle=140)
plt.title('Project Complexity')
plt.axis('equal')
plt.show()
```



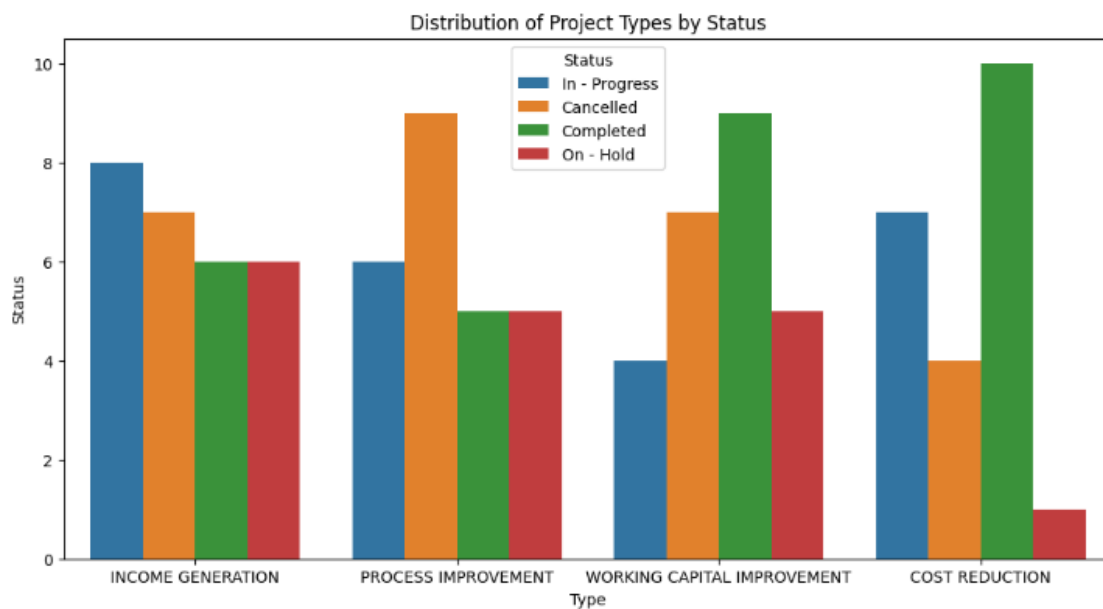
```
plt.figure(figsize=(8,8))
plt.pie(pha, labels=pha.index,autopct='%1.1f%%', startangle=140)
plt.title('Project Phase')
plt.axis('equal')
plt.show()
```



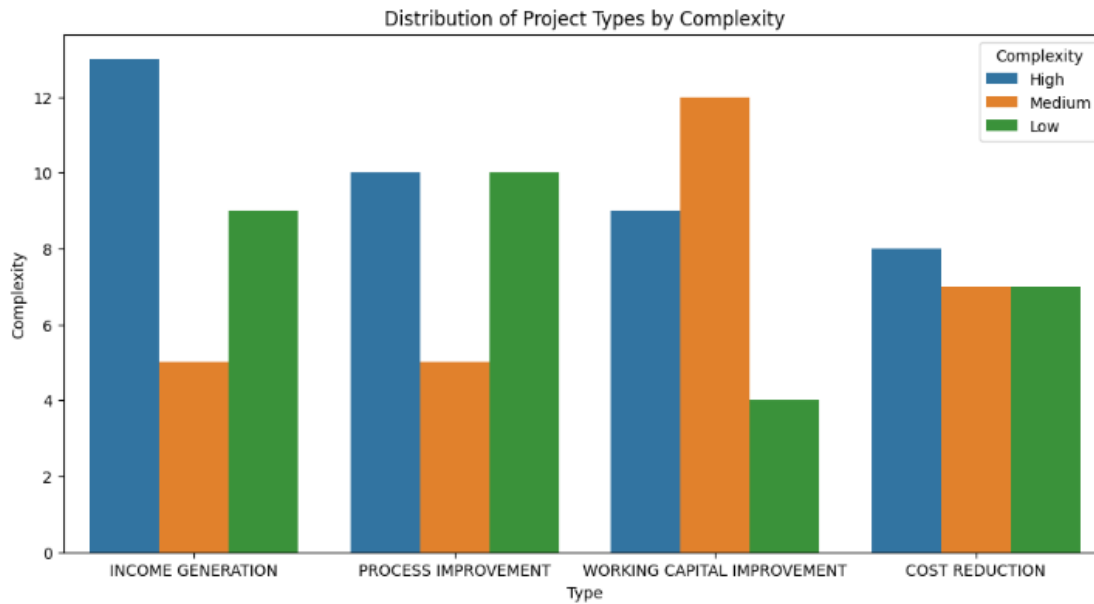

```
> plt.figure(figsize=(12,6))
sns.countplot(data=df, x='Project Type', hue='Department')
plt.xlabel('Project Type')
plt.ylabel('Department')
plt.title('Project Type by Department')
plt.show()
```



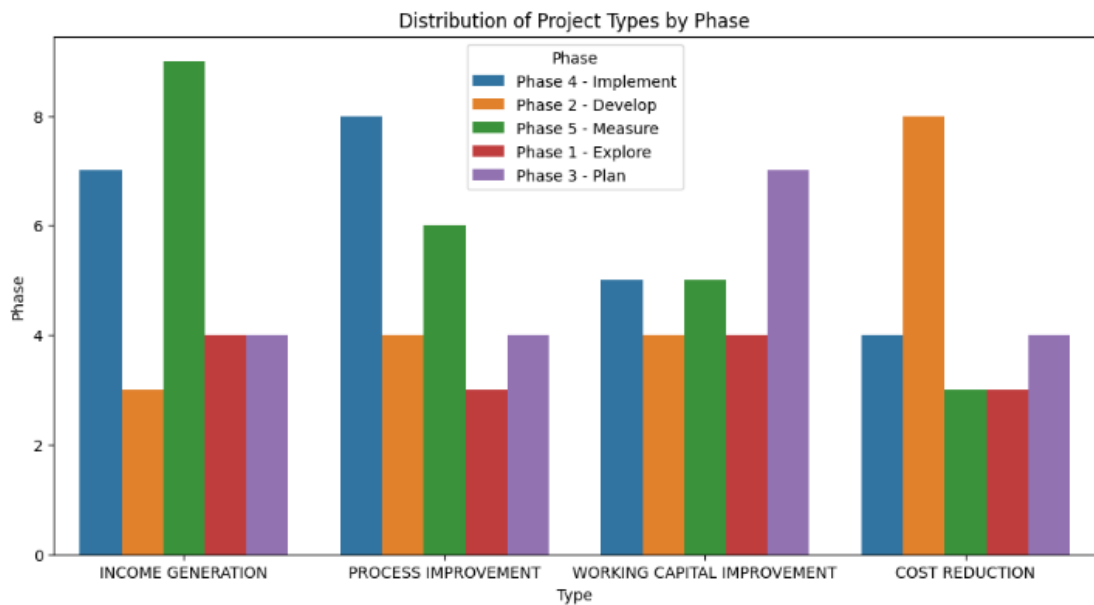
```
plt.figure(figsize=(12,6))
sns.countplot(data=df, x='Project Type', hue='Status')
plt.xlabel('Type')
plt.ylabel('Status')
plt.title('Distribution of Project Types by Status')
plt.show()
```



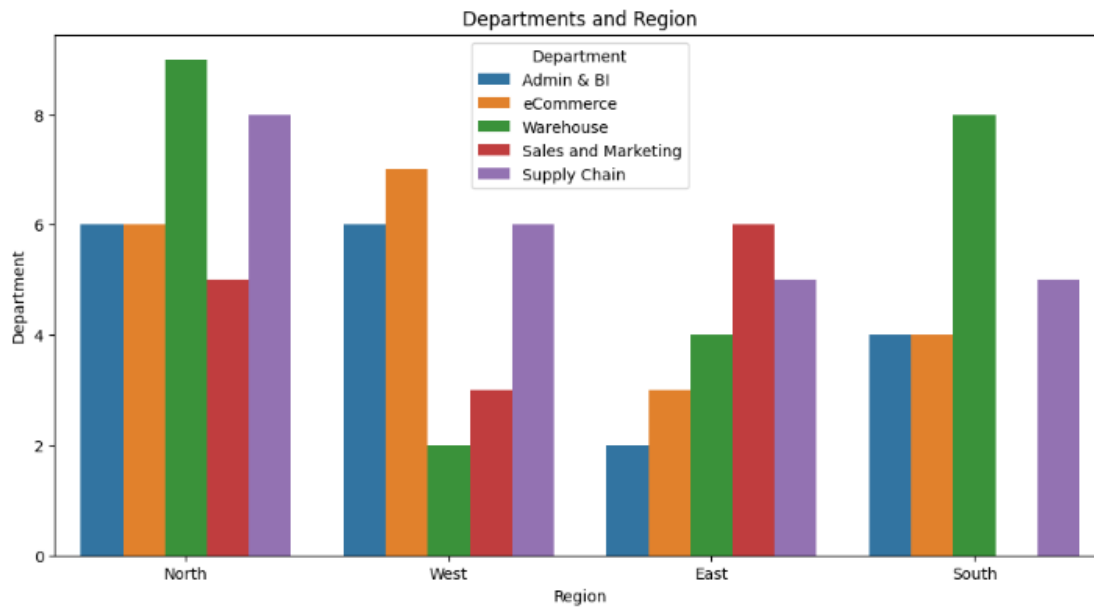
```
plt.figure(figsize=(12,6))
sns.countplot(data=df, x='Project Type', hue='Complexity')
plt.xlabel('Type')
plt.ylabel('Complexity')
plt.title('Distribution of Project Types by Complexity')
plt.show()
```



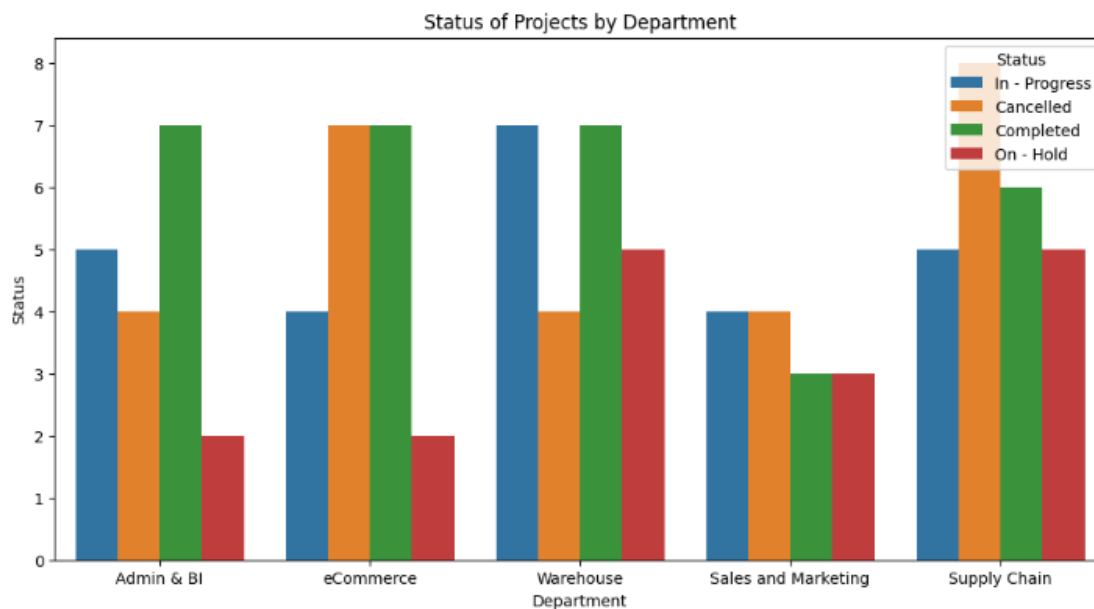
```
plt.figure(figsize=(12,6))
sns.countplot(data=df, x='Project Type', hue='Phase')
plt.xlabel('Type')
plt.ylabel('Phase')
plt.title('Distribution of Project Types by Phase')
plt.show()
```



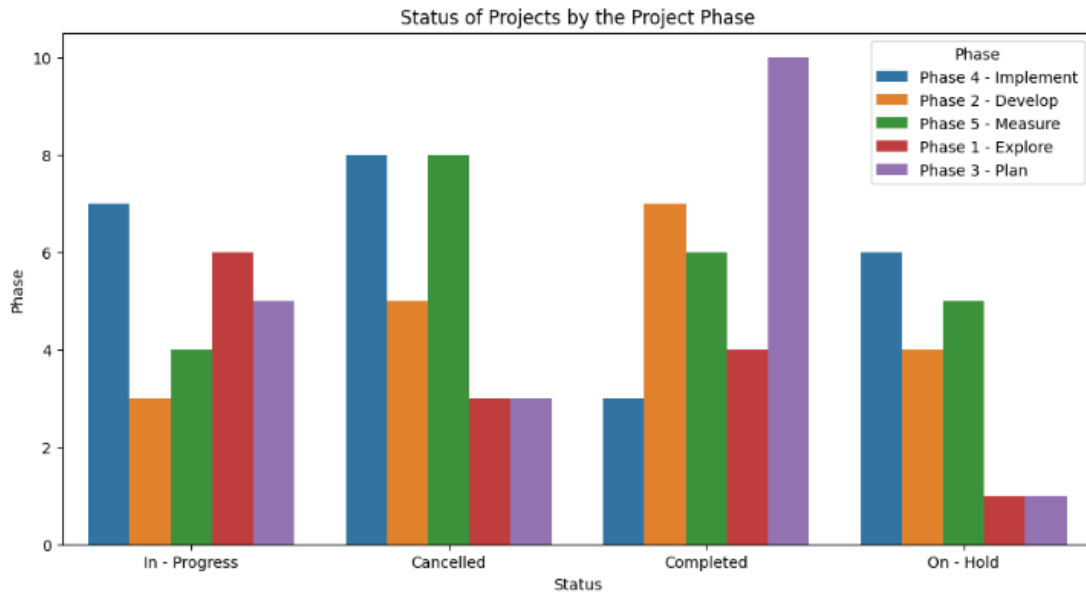
```
plt.figure(figsize=(12,6))
sns.countplot(data=df, x='Region', hue='Department')
plt.xlabel('Region')
plt.ylabel('Department')
plt.title('Departments and Region')
plt.show()
```



```
plt.figure(figsize=(12,6))
sns.countplot(data=df, x='Department', hue='Status')
plt.xlabel('Department')
plt.ylabel('Status')
plt.title('Status of Projects by Department')
plt.show()
```



```
plt.figure(figsize=(12,6))
sns.countplot(data=df, x='Status', hue='Phase')
plt.xlabel('Status')
plt.ylabel('Phase')
plt.title('Status of Projects by the Project Phase')
plt.show()
```



```
plt.figure(figsize=(12,6))
sns.countplot(data=df, x='Status', hue='Complexity')
plt.xlabel('Status')
plt.ylabel('Complexity')
plt.title('Status of Projects by Complexity')
plt.show()
```

