

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

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

Import CSV File

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

Data Preprocessing

.head()

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

```
df.head()

{"summary":{"\n  \"name\": \"df\",\n  \"rows\": 99,\n  \"fields\": [\n    {\n      \"column\": \"Project Name\",\n      \"properties\": {\n        \"dtype\": \"string\",\n        \"num_unique_values\": 99,\n        \"samples\": [\n          \"Next Gala\",\n          \"The Coding Expert\",\n          \"Debug Entity\"],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Project Description\",\n      \"properties\": {\n        \"dtype\": \"string\",\n        \"num_unique_values\": 95,\n        \"samples\": [\n          \"Uses Continuous Continuous Learning Algorithms To Constantly Improve The Recommendations; We Constantly See Product Changes And Improvements. The More You Use Our Platform, The More We Learn Which Leads To Our Clients Benefiting More And More.\",\n          \"If You Are Making Less Than $50,000 With Your Startup Company, You\u2019Re In Luck.\"
```

Companies Who Are Making Less Than That Are Eligible For The Associations Now Program. Associations Now Is An Algorithm That Helps Startups Find Niche Or Narrow Topics That Are Related To Your Industry Or Target Audience. It Is Like Word Of Mouth Marketing, But You Pay For It!\"
\"In A Saturated Market Of Founders Turning To Content Marketing, Finding A Way To Stand Out From The Crowd Is A Competitive Advantage. \"2019S Ai Takes A Deep Dive Into A Topic\"2019S Domain And Produces A Content Strategy In Minutes In A Standalone Ai Tool. Rapid, Accurate, And Tailored Content Is A Step Towards Building A Meaningful Audience That Can Amplify Your Voice.\"

```

],\n      {\n        \"semantic_type\": \"\",\n        \"description\": \"\",\n        \"column\": \"Project Type\",\n        \"properties\": {\n          \"dtype\": \"category\",\n          \"num_unique_values\": 4,\n          \"samples\": [\n            \"PROCESS IMPROVEMENT\",\n            \"COST REDUCTION\",\n            \"INCOME GENERATION\"\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\",\n          \"column\": \"Project Manager\",\n          \"properties\": {\n            \"dtype\": \"category\",\n            \"num_unique_values\": 7,\n            \"samples\": [\n              \"Yael Wilcox\",\n              \"Brenda Chandler\",\n              \"Aleena Khan\"\n            ],\n            \"semantic_type\": \"\",\n            \"description\": \"\",\n            \"column\": \"Region\",\n            \"properties\": {\n              \"dtype\": \"category\",\n              \"num_unique_values\": 4,\n              \"samples\": [\n                \"West\",\n                \"South\",\n                \"North\"\n              ],\n              \"semantic_type\": \"\",\n              \"description\": \"\",\n              \"column\": \"Department\",\n              \"properties\": {\n                \"dtype\": \"category\",\n                \"num_unique_values\": 5,\n                \"samples\": [\n                  \"eCommerce\",\n                  \"Supply Chain\",\n                  \"Warehouse\"\n                ],\n                \"semantic_type\": \"\",\n                \"description\": \"\",\n                \"column\": \"Project Cost\",\n                \"properties\": {\n                  \"dtype\": \"number\",\n                  \"std\": 1076543,\n                  \"min\": 2418301,\n                  \"max\": 5974815,\n                  \"num_unique_values\": 99,\n                  \"samples\": [\n                    3261249,\n                    3978102,\n                    4790417\n                  ],\n                  \"semantic_type\": \"\",\n                  \"description\": \"\",\n                  \"column\": \"Project Benefit\",\n                  \"properties\": {\n                    \"dtype\": \"number\",\n                    \"std\": 216401,\n                    \"min\": 8422578,\n                    \"max\": 9165877,\n                    \"num_unique_values\": 99,\n                    \"samples\": [\n                      8696481,\n                      8488880,\n                      8872443\n                    ],\n                    \"semantic_type\": \"\",\n                    \"description\": \"\",\n                    \"column\": \"Complexity\",\n                    \"properties\": {\n                      \"dtype\": \"category\",\n                      \"num_unique_values\": 3,\n                      \"samples\": [\n                        \"High\",\n                        \"Medium\",\n                        \"Low\"\n                      ],\n                      \"semantic_type\": \"\",\n                      \"description\": \"\",\n                      \"column\": \"Status\",\n                      \"properties\": {\n

```

```
{\n      \"dtype\": \"category\", \n      \"num_unique_values\": 4, \n      \"samples\": [\n        \"Cancelled\", \n        \"0n - Hold\", \n        \"In - Progress\" \n      ], \n      \"semantic_type\": \"\", \n      \"description\": \"\" \n    }, \n    {\n      \"column\": \"Completion%\", \n      \"properties\": {\n        \"dtype\": \"category\", \n        \"num_unique_values\": 22, \n        \"samples\": [\n          \"77%\", \n          \"86%\", \n          \"72%\" \n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\" \n      }, \n      {\n        \"column\": \"Phase\", \n        \"properties\": {\n          \"dtype\": \"category\", \n          \"num_unique_values\": 5, \n          \"samples\": [\n            \"Phase 2 - Develop\", \n            \"Phase 3 - Plan\", \n            \"Phase 5 - Measure\" \n          ], \n          \"semantic_type\": \"\", \n          \"description\": \"\" \n        }, \n        {\n          \"column\": \"Year\", \n          \"properties\": {\n            \"dtype\": \"number\", \n            \"std\": 1, \n            \"min\": 2021, \n            \"max\": 2025, \n            \"num_unique_values\": 5, \n            \"samples\": [\n              2022, \n              2025, \n              2023 \n            ], \n            \"semantic_type\": \"\", \n            \"description\": \"\" \n          }, \n          {\n            \"column\": \"Month\", \n            \"properties\": {\n              \"dtype\": \"number\", \n              \"std\": 3, \n              \"min\": 1, \n              \"max\": 12, \n              \"num_unique_values\": 12, \n              \"samples\": [\n                12, \n                11, \n                2 \n              ], \n              \"semantic_type\": \"\", \n              \"description\": \"\" \n            }, \n            {\n              \"column\": \"Start Date\", \n              \"properties\": {\n                \"dtype\": \"object\", \n                \"num_unique_values\": 49, \n                \"samples\": [\n                  \"3/1/2022\", \n                  \"8/1/2025\", \n                  \"11/1/2025\" \n                ], \n                \"semantic_type\": \"\", \n                \"description\": \"\" \n              }, \n              {\n                \"column\": \"End Date\", \n                \"properties\": {\n                  \"dtype\": \"object\", \n                  \"num_unique_values\": 43, \n                  \"samples\": [\n                    \"8/1/2025\", \n                    \"12/1/2023\", \n                    \"2/1/2024\" \n                  ], \n                  \"semantic_type\": \"\", \n                  \"description\": \"\" \n                } \n              } \n            } \n          ], \n          \"type\": \"dataframe\", \"variable_name\": \"df\"}
```

.tail()

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

```
df.tail()

{"repr_error": "0", "type": "dataframe"}
```

.shape

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

```
df.shape
```

```
(99, 16)
```

.columns

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

```
df.columns  
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  
Project Name      object  
Project Description  object  
Project Type      object  
Project Manager    object  
Region            object  
Department         object  
Project Cost      int64  
Project Benefit    int64  
Complexity        object  
Status            object  
Completion%       object  
Phase            object  
Year             int64  
Month            int64  
Start Date       object  
End Date         object  
dtype: 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)
```

.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
```

.describe()

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

```
df.describe()

{"summary":{"\n  \"name\": \"df\",\n  \"rows\": 8,\n  \"fields\": [\n    {\n      \"column\": \" Project Cost \",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 2012891.024302758,\n        \"min\": 99.0,\n        \"max\": 5974815.0,\n        \"num_unique_values\": 8,\n        \"samples\": [\n          4156649.3636363638,\n          4172827.0,\n          99.0\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \" Project Benefit \",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 4040686.378056789,\n        \"min\": 99.0,\n        \"max\": 9165877.0,\n        \"num_unique_values\": 8,\n        \"samples\": [\n          8828178.131313132,\n          8846243.0,\n          99.0\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Year\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 913.5051225365318,\n        \"min\": 1.4022098764525621,\n        \"max\": 2025.0,\n        \"num_unique_values\": 7,\n        \"samples\": [\n          99.0,\n          2022.7474747474748,\n          2024.0\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Month\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 32.92770843152441,\n        \"min\": 1.0,\n        \"max\":
```

```
99.0,\n          \"num_unique_values\": 8,\n          \"samples\": [\n7.151515151515151,\n          7.0,\n          99.0\n],\n\"semantic_type\": \"\",,\n\"description\": \"\"\n}\n}\n]\n}","type":"dataframe"}
```

.value_counts()

Returns the number of all unique values in a column.

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

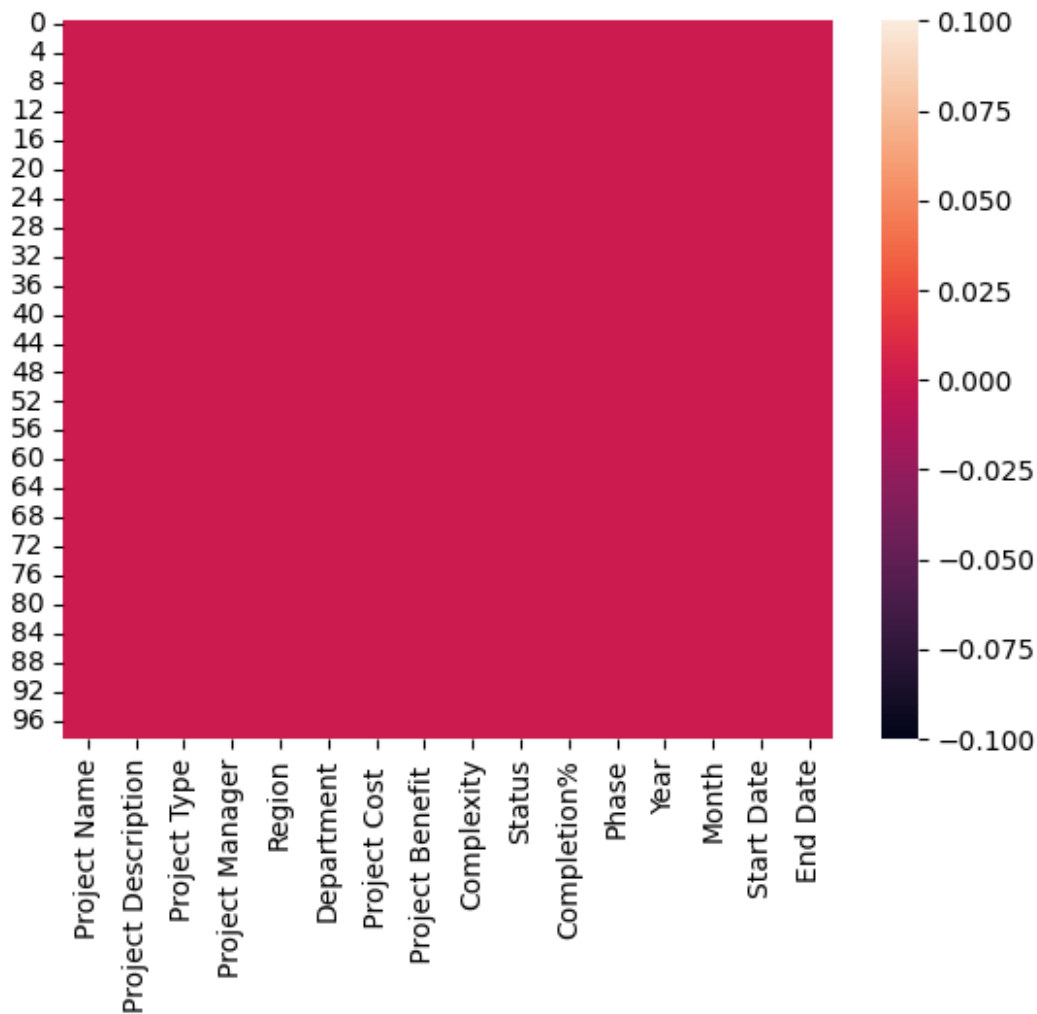
```
Project Type
INCOME GENERATION      27
PROCESS IMPROVEMENT    25
WORKING CAPITAL IMPROVEMENT  25
COST REDUCTION          22
Name: count, dtype: int64
```

.isnull()

Checks for null values.

```
df.isnull()
```

```
{"summary":{"\n  \"name\": \"df\",\n  \"rows\": 99,\n  \"fields\": [\n    {\n      \"column\": \"Project Name\",\n      \"properties\": {\n        \"dtype\": \"boolean\",\n        \"num_unique_values\": 1,\n        \"samples\": [\n          false\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Project Description\",\n      \"properties\": {\n        \"dtype\": \"boolean\",\n        \"num_unique_values\": 1,\n        \"samples\": [\n          false\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Project Type\",\n      \"properties\": {\n        \"dtype\": \"boolean\",\n        \"num_unique_values\": 1,\n        \"samples\": [\n          false\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Project Manager\",\n      \"properties\": {\n        \"dtype\": \"boolean\",\n        \"num_unique_values\": 1,\n        \"samples\": [\n          false\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Region\",\n      \"properties\": {\n        \"dtype\": \"boolean\",\n        \"num_unique_values\": 1,\n        \"samples\": [\n          false\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Department\",\n      \"properties\": {\n        \"dtype\": \"boolean\",\n        \"num_unique_values\": 1,\n        \"samples\": [\n          false\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Project Cost \"
```

Data Analysis

```
df["Start Date"] = pd.to_datetime(df["Start Date"])
df["End Date"] = pd.to_datetime(df["End Date"])

df['Start Month'] = df['Start Date'].dt.month_name()
df['End Month'] = df['End Date'].dt.month_name()
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...

	Project Type	Project Manager	Region	
Department \				
0	INCOME GENERATION	Yael Wilcox	North	
Admin & BI				
1	INCOME GENERATION	Brenda Chandler	West	
eCommerce				
2	INCOME GENERATION	Nyasia Hunter	North	
Warehouse				
3	PROCESS IMPROVEMENT	Brenda Chandler	East	Sales and
Marketing				
4	WORKING CAPITAL IMPROVEMENT	Jaylyn Mckenzie	East	
eCommerce				
..	
...				
94	WORKING CAPITAL IMPROVEMENT	Nyasia Hunter	South	Supply
Chain				
95	INCOME GENERATION	Kamari Norris	North	
Warehouse				
96	PROCESS IMPROVEMENT	Yael Wilcox	West	Supply
Chain				
97	COST REDUCTION	Jaylyn Mckenzie	East	
eCommerce				
98	WORKING CAPITAL IMPROVEMENT	Nyasia Hunter	West	Sales and
Marketing				

	Project Cost	Project Benefit	Complexity	Status
Completion% \				
0	3648615	8443980	High	In - Progress
77%				
1	4018835	9012225	High	Cancelled
80%				
2	4285483	9078339	High	Completed

100%				
3	5285864	8719006	High	Cancelled
75%				
4	5785601	8630148	High	Completed
100%				
..
...				
94	5259436	8817917	Medium	On - Hold
80%				
95	4790417	8872443	Medium	In - Progress
73%				
96	4283481	8895152	Low	Completed
100%				
97	4606575	8658343	High	In - Progress
77%				
98	5054482	8422578	High	In - Progress
83%				

	Phase	Year	Month	Start Date	End Date	Start Month
\						
0	Phase 4 - Implement	2021	2	2021-02-01	2021-06-01	February
1	Phase 2 - Develop	2021	3	2021-03-01	2021-06-01	March
2	Phase 4 - Implement	2021	3	2021-03-01	2021-06-01	March
3	Phase 5 - Measure	2021	3	2021-03-01	2021-06-01	March
4	Phase 1 - Explore	2021	3	2021-03-01	2021-06-01	March
..
94	Phase 2 - Develop	2025	8	2025-08-01	2025-11-01	August
95	Phase 4 - Implement	2025	9	2025-09-01	2025-12-01	September
96	Phase 3 - Plan	2025	11	2025-11-01	2026-03-01	November
97	Phase 4 - Implement	2025	11	2025-11-01	2026-03-01	November
98	Phase 3 - Plan	2025	12	2025-12-01	2026-03-01	December

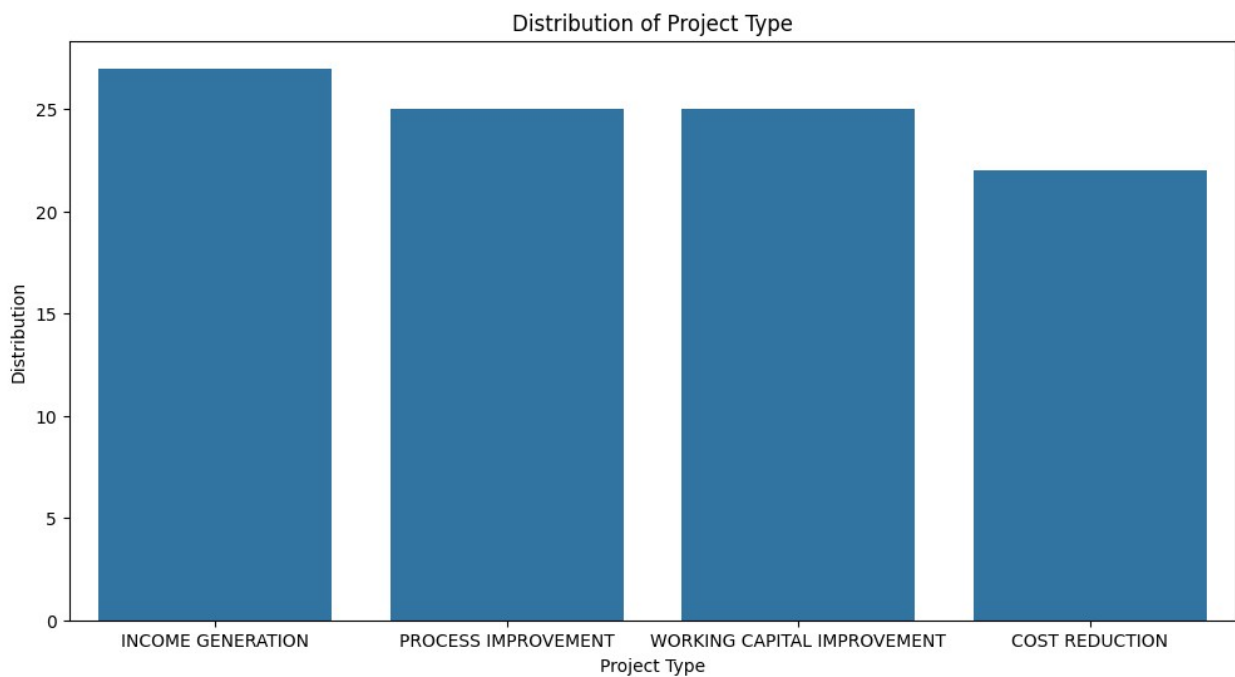
	End Month
0	June
1	June
2	June
3	June
4	June
..	...

```
94 November
95 December
96 March
97 March
98 March
```

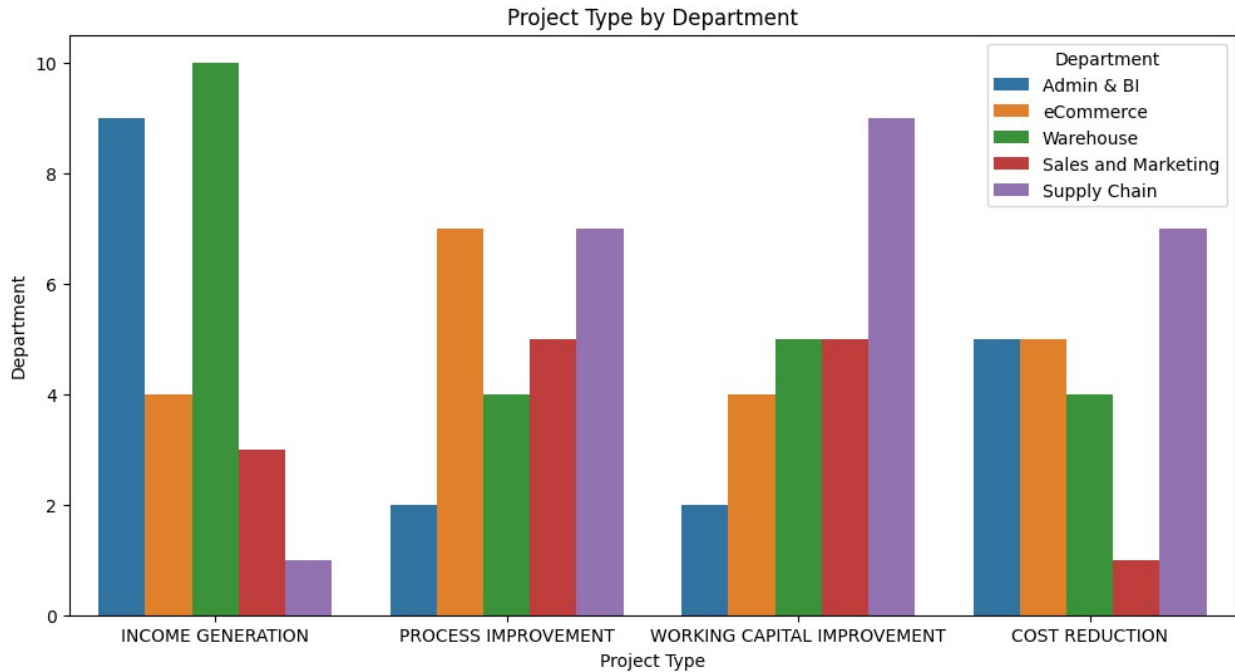
```
[99 rows x 18 columns]
```

Data Visualization

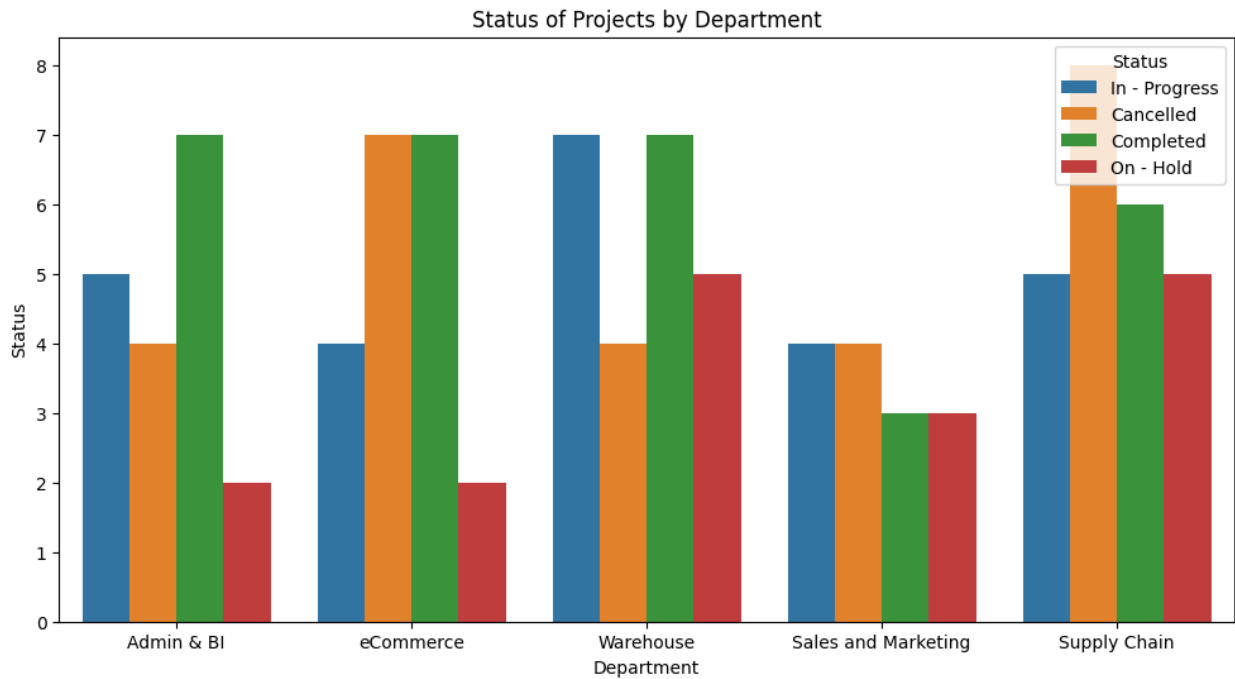
```
plt.figure(figsize=(12,6))
sns.countplot(data=df, x='Project Type')
plt.xlabel('Project Type')
plt.ylabel('Distribution')
plt.title('Distribution of Project Type')
plt.show()
```



```
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()
```

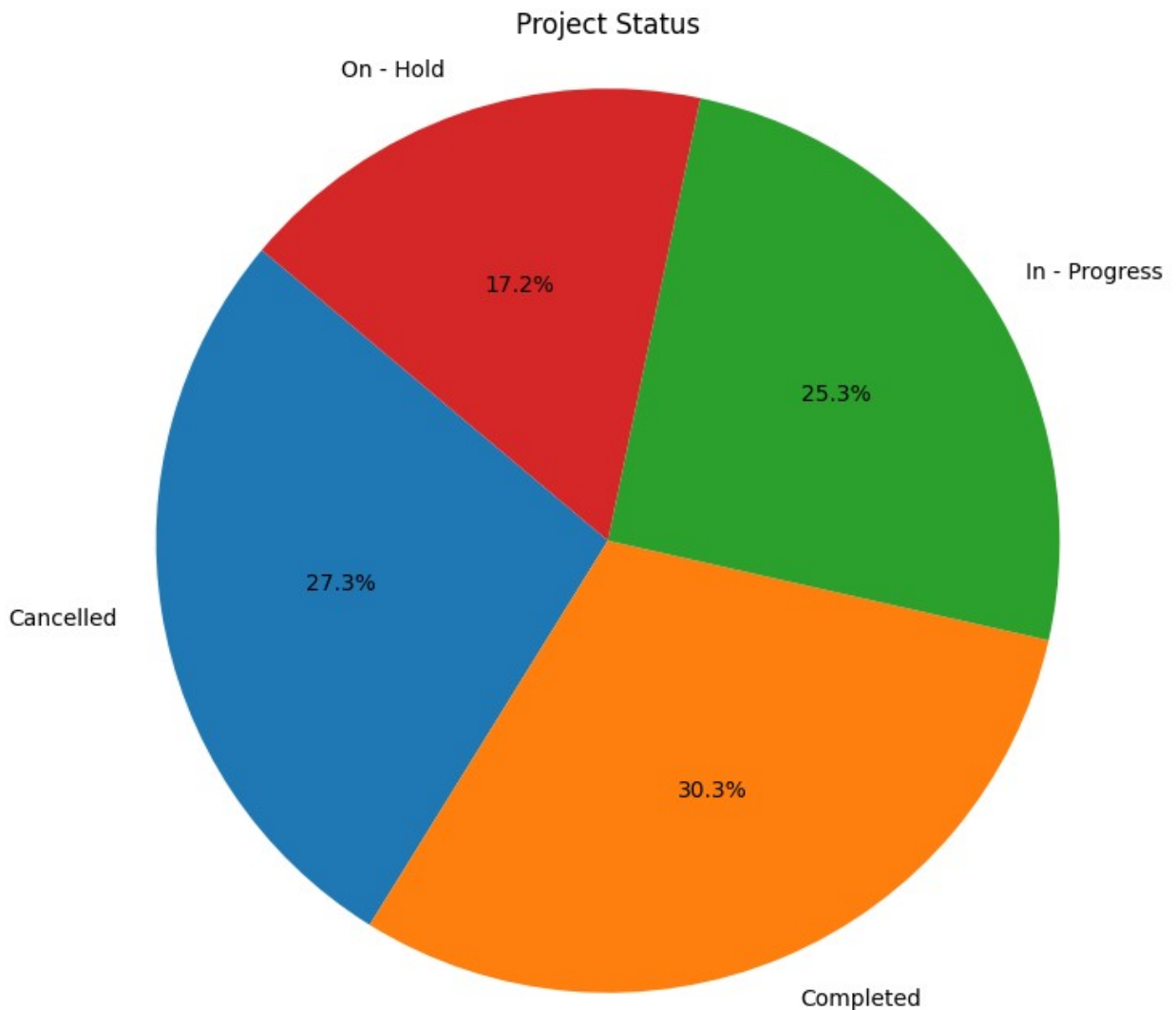


```
plt.figure(figsize=(12,6))
sns.countplot(data=df, x='Department', hue='Status')
plt.xlabel('Department')
plt.ylabel('Status')
plt.title('Status of Projects by Department')
plt.show()
```

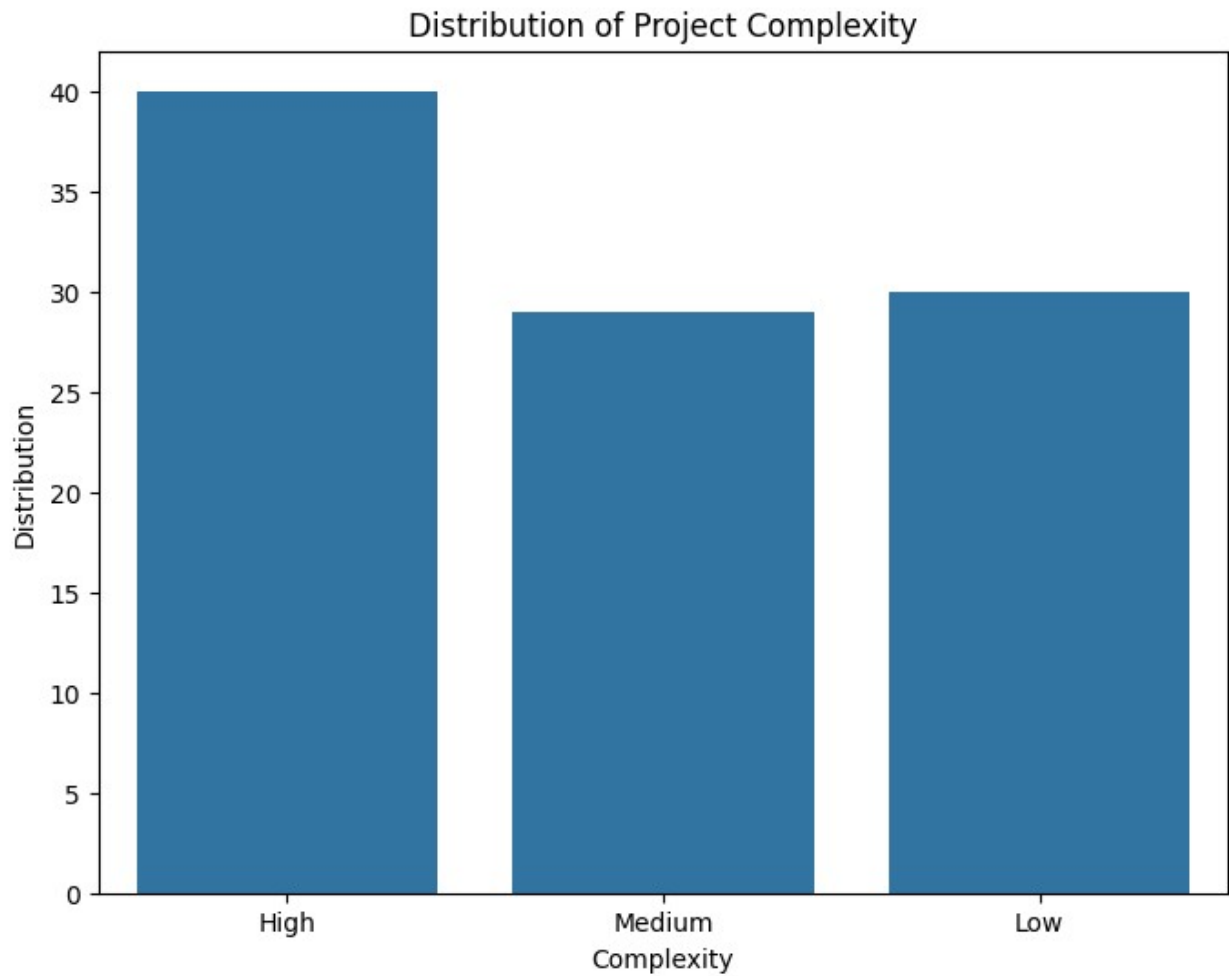


```
stats = df.groupby('Status').size()

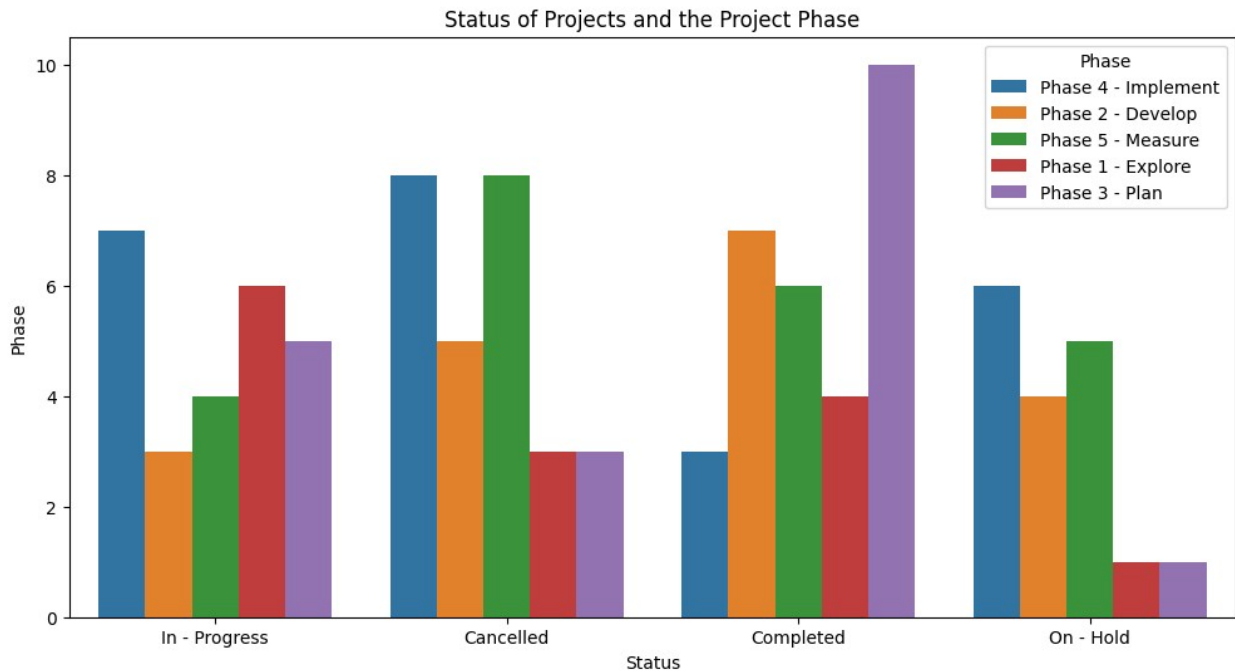
plt.figure(figsize=(8,8))
plt.pie(stats, labels=stats.index, autopct='%1.1f%%', startangle=140)
plt.title('Project Status')
plt.axis('equal')
plt.show()
```



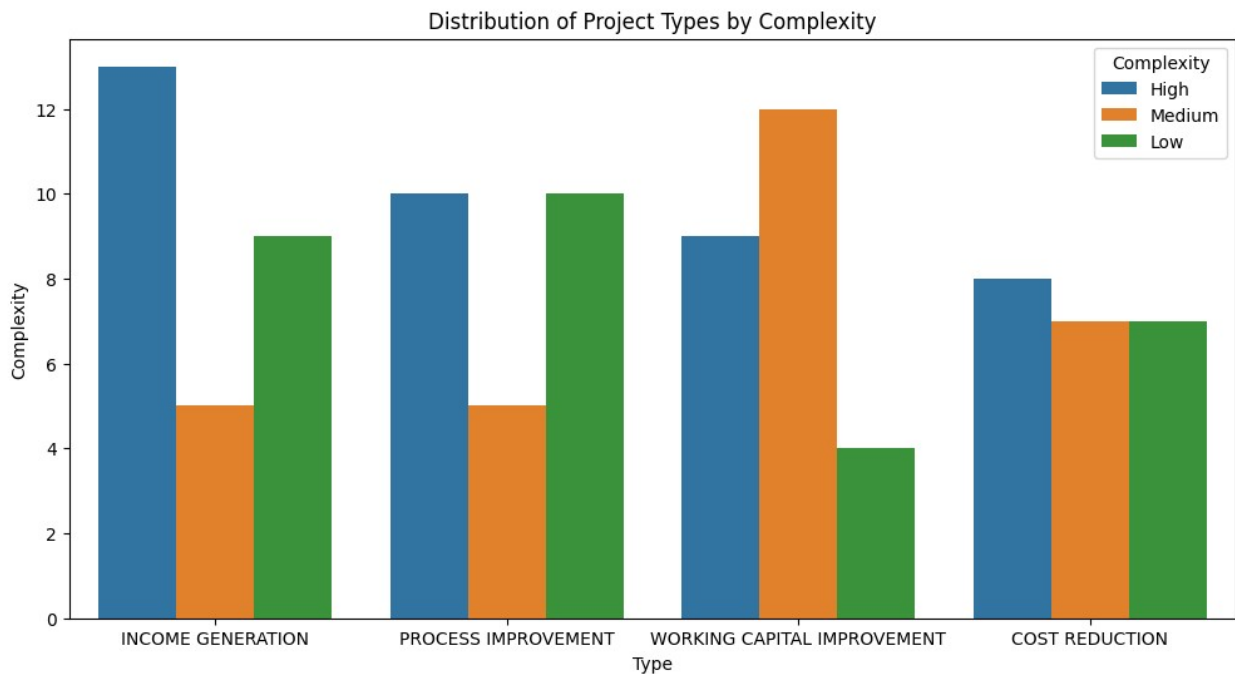
```
plt.figure(figsize=(8,6))
sns.countplot(data=df, x='Complexity')
plt.xlabel('Complexity')
plt.ylabel('Distribution')
plt.title('Distribution of Project Complexity')
plt.show()
```



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

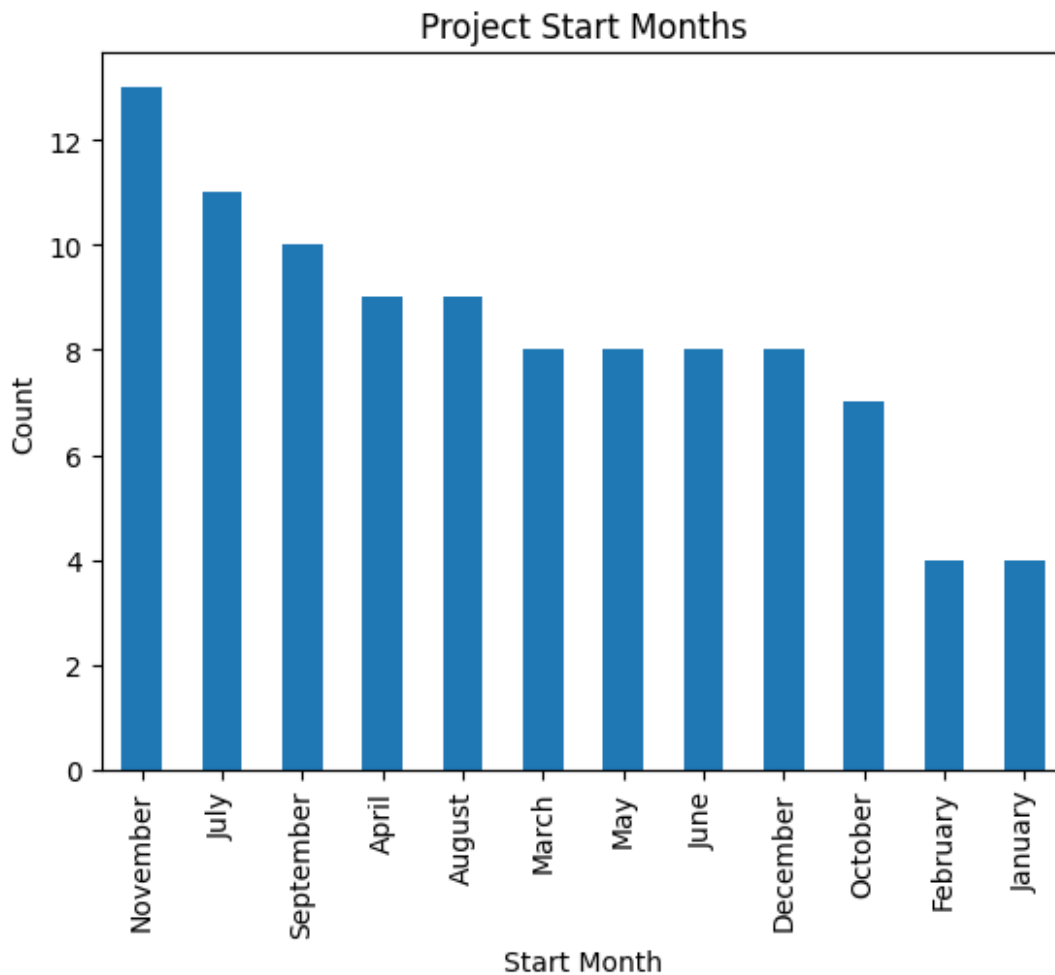


```
plt.figure(figsize=(12,6))
sns.countplot(data=df, x='Project Type', hue='Complexity')
plt.xlabel('Type')
plt.ylabel('Complexity')
plt.title('Distribution of Project Types by Complexity')
plt.show()
```



```
Start_Month = df['Start Month'].value_counts()
End_Month = df['End Month'].value_counts()

Start_Month.plot(kind='bar')
plt.xlabel('Start Month')
plt.ylabel('Count')
plt.title('Project Start Months')
plt.show()
```



```
End_Month.plot(kind='bar')
plt.xlabel('End Month')
plt.ylabel('Count')
plt.title('Project End Months')
plt.show()
```