Jan Jacob

﻿import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

df = pd.read\_csv("got\_battles.csv")

1. Are there any missing values? If yes, drop the rows that have missing values.

df.info()

**# There are missing values**

df=df.dropna()

1. Which year had the most battles?

![Chart, bar chart

Description automatically generated]()

sns.countplot(x="year", data=df)

**# Year 299 had the most battles (17)**

1. Which year had the most wins from attackers perspective?

![Chart, bar chart

Description automatically generated]()

wins=df.loc[df["attacker\_outcome"]=="win"]

wins\_by\_year = wins.groupby('year').count()["name"]

plt.bar(wins\_by\_year.index,wins\_by\_year)

**# Year 299 had the most wins (14)**

1. Which year had the least battles?

![Chart, bar chart

Description automatically generated]()

sns.countplot(x="year", data=df)

**# Year 298 had the least battles (7)**

1. Which region had the most battles?

![Chart, bar chart

Description automatically generated]()

sns.countplot(x="region", data=df)

plt.xticks(rotation=90)

**# Riverdale had the most battles (13)**

1. Which region had least battles?

![Chart, bar chart

Description automatically generated]()

sns.countplot(x="region", data=df)

plt.xticks(rotation=90)

**# Beyond the Wall had the least battles (1)**

1. What was the outcome of all battles from attacker's perspective?

![Chart, bar chart

Description automatically generated]()

sns.countplot(x="attacker\_outcome", data=df)

**# The attackers had 5 losses and 28 wins in total**

1. How common were different types of battles?

![Chart, bar chart

Description automatically generated]()

sns.countplot(x="battle\_type", data=df)

**# ambush - 10, pitched battle - 11, razing - 1, siege - 11**

1. Which king was the most attacked?

![Chart, bar chart

Description automatically generated]()

sns.countplot(x="defender\_king", data=df)

plt.xticks(rotation=90)

**# Joffrey/Tommen Baratheon was attacked the most (12)**

1. Which king attacked the most?

![Chart, bar chart

Description automatically generated]()

sns.countplot(x="attacker\_king", data=df)

plt.xticks(rotation=90)

**# Joffrey/Tommen Baratheon attacked the most (12)**

1. Which year had highest chances of winning from attacker's perspective?

![Chart, bar chart

Description automatically generated]()

﻿percent\_winning\_by\_year = wins.groupby('year').count()["name"] / df.groupby("year").count()["name"]

plt.bar(percent\_winning\_by\_year.index,percent\_winning\_by\_year)

plt.xticks(rotation=90)

**# year 300 has the highest win percentage of winning**

1. Compare the mean deaths for different years?

![Chart, bar chart

Description automatically generated]()

mean\_deaths\_by\_year = df.groupby(by="year").mean()["total deaths"]

plt.bar(mean\_deaths\_by\_year.index,mean\_deaths\_by\_year)

**# 300 has the lowest mean with 643.222222, then 298 has a mean of 734.714286, and 299 has a much higher mean of 1205.176471**

1. Compare the mean deaths for different battle types?

![Chart, bar chart

Description automatically generated]()

mean\_deaths\_by\_battle\_type = df.groupby(by="battle\_type").mean()["total deaths"]

plt.bar(mean\_deaths\_by\_battle\_type.index,mean\_deaths\_by\_battle\_type)

**# pitched battle has the lowest mean of 549.545455, then siege has a mean of 1098.818182,**

**# then razing has a mean of 1145, and ambush has the highest mean of 1214.3**

1. Plot the total(sum of) deaths for different battle types?

![Chart, bar chart

Description automatically generated]()

sns.countplot(x="battle\_type", data=df)

1. Using visualizations share any three interesting insights (other than the above questions) from this dataset.

tc=df.corr()

sns.heatmap(tc)

![Square

Description automatically generated]()

**## This shows a strong correlation between battle\_number and the year**

sns.countplot(x="major\_death", data=df)

![Chart, bar chart

Description automatically generated]()

**# this shows that more battles did not have a major death (20 compared to 12)**

sns.countplot(x="summer", data=df)

![Chart, bar chart

Description automatically generated]()

**# this shows that more battles happened in the summer (24 compared to 8)**