```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from wordcloud import WordCloud
from wordcloud import STOPWORDS
from matplotlib.pyplot import figure
import seaborn as sns
%matplotlib inline
```

```
df = pd.read_csv("/content/neo_data_year.csv")
```

```
df.head()
```

	Date	ID	Name	Neo Reference ID	NASA JPL URL	Absol Magnit
0	2023-02-13	2005879	5879 Almeria (1992 CH1)	2005879	https://ssd.jpl.nasa.gov/tools/sbdb_lookup.htm...	17
1	2023-02-13	2138911	138911 (2001 AE2)	2138911	https://ssd.jpl.nasa.gov/tools/sbdb_lookup.htm...	19
2	2023-02-13	2187026	187026 (2005 EK70)	2187026	https://ssd.jpl.nasa.gov/tools/sbdb_lookup.htm...	17
3	2023-02-13	2380818	380818 (2005 YV128)	2380818	https://ssd.jpl.nasa.gov/tools/sbdb_lookup.htm...	20
4	2023-02-13	2518810	518810 (2010 CF19)	2518810	https://ssd.jpl.nasa.gov/tools/sbdb_lookup.htm...	21

```
ndf=df.dropna()
ndf.info()
```

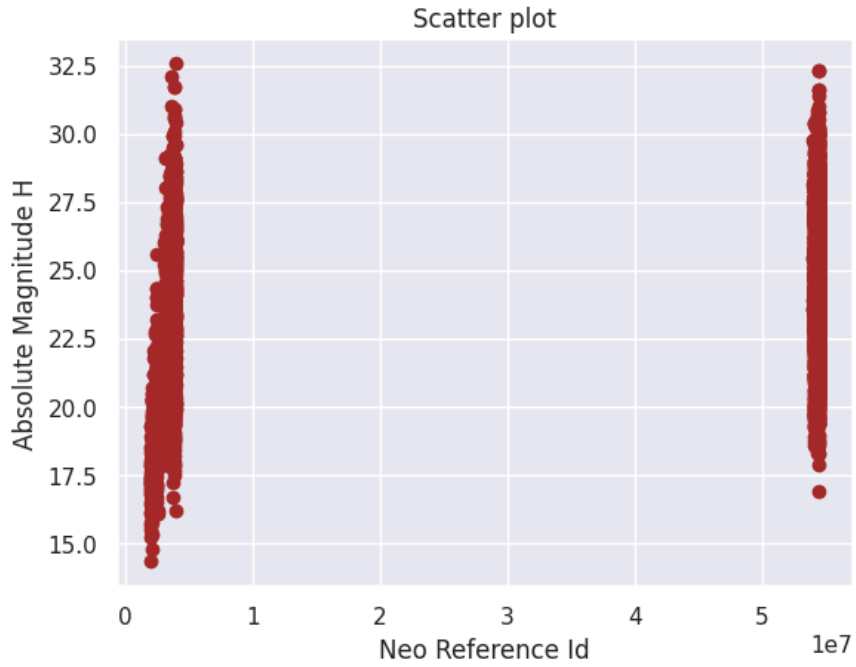
```
<class 'pandas.core.frame.DataFrame'>
Index: 7070 entries, 0 to 7070
Data columns (total 14 columns):
 #   Column                             Non-Null Count  Dtype
---  ------                             --------------  -----
 0   Date                               7070 non-null   object
 1   ID                                 7070 non-null   int64
 2   Name                               7070 non-null   object
 3   Neo Reference ID                   7070 non-null   int64
 4   NASA JPL URL                       7070 non-null   object
 5   Absolute Magnitude H               7070 non-null   float64
 6   Estimated Diameter Min (km)        7070 non-null   float64
 7   Estimated Diameter Max (km)        7070 non-null   float64
 8   Is Potentially Hazardous Asteroid  7070 non-null   bool
 9   Close Approach Date Full           7070 non-null   object
 10  Relative Velocity (km/h)           7070 non-null   float64
 11  Miss Distance (astronomical units) 7070 non-null   float64
 12  Miss Distance (kilometers)         7070 non-null   float64
 13  Orbiting Body                      7070 non-null   object
dtypes: bool(1), float64(6), int64(2), object(5)
memory usage: 780.2+ KB
```

```
sns.set()
```

SCATTER PLOT

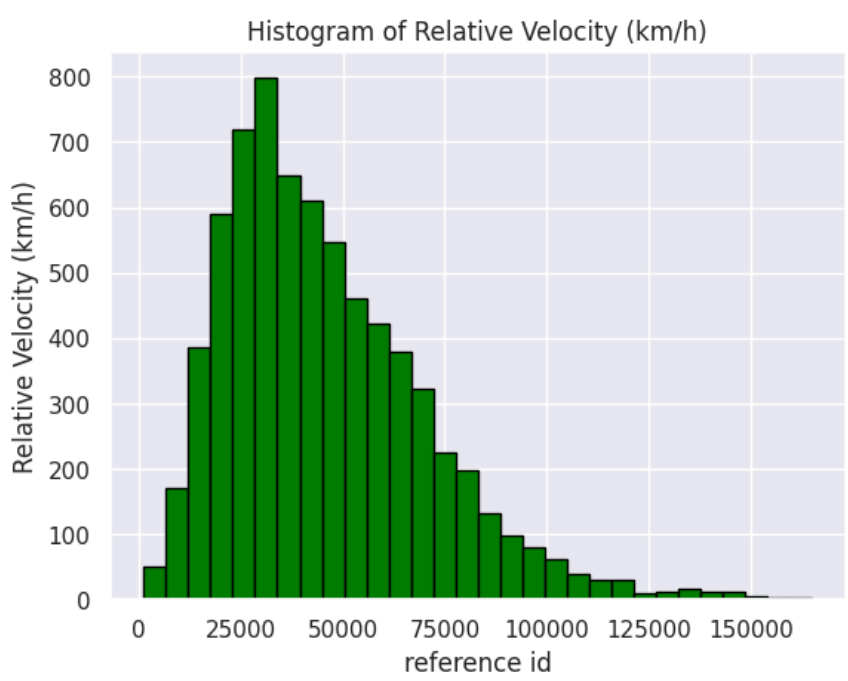
```
plt.scatter(df['Neo Reference ID'], df['Absolute Magnitude H'], c='brown')
plt.title('Scatter plot')
plt.xlabel('Neo Reference Id')
plt.ylabel('Absolute Magnitude H')
```

```
plt.show()
```



HISTOGRAM

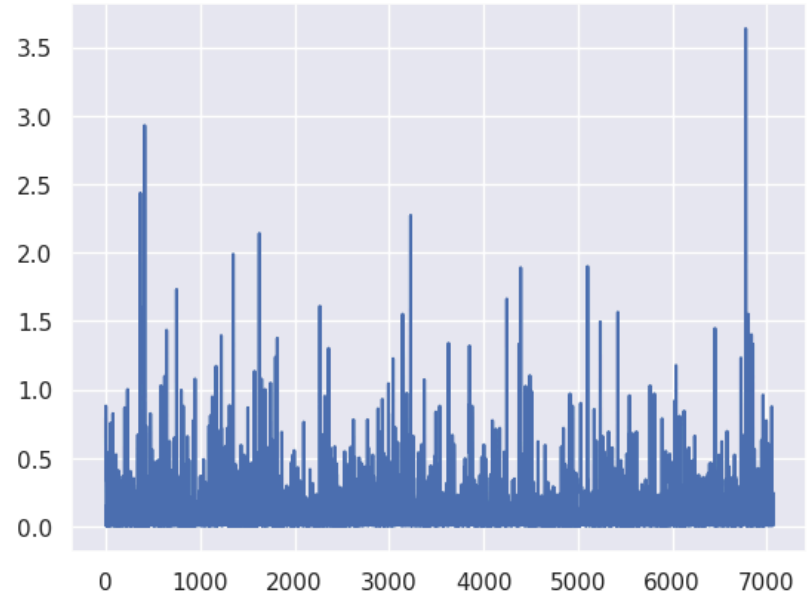
```
plt.hist(df['Relative Velocity (km/h)'],color='green',edgecolor='black',bins=30)
plt.title('Histogram of Relative Velocity (km/h)')
plt.xlabel('reference id')
plt.ylabel('Relative Velocity (km/h)')
plt.show()
```



LINEPLOT

```
df['Estimated Diameter Min (km)'].plot()
```

```
<Axes: >
```




```
pip install wordcloud matplotlib
```

```
Requirement already satisfied: wordcloud in /usr/local/lib/python3.10/dist-packages (1.9.3)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-packages (3.7.1)
Requirement already satisfied: numpy>=1.6.1 in /usr/local/lib/python3.10/dist-packages (from wordcloud) (1.25.2)
Requirement already satisfied: pillow in /usr/local/lib/python3.10/dist-packages (from wordcloud) (9.4.0)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (1.2.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (4.51.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (1.4.5)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (24.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib) (2.8
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->matplot
```

WORDCLOUD

```
text= " ".join(item for item in ndf['Name'])
print(text)
```

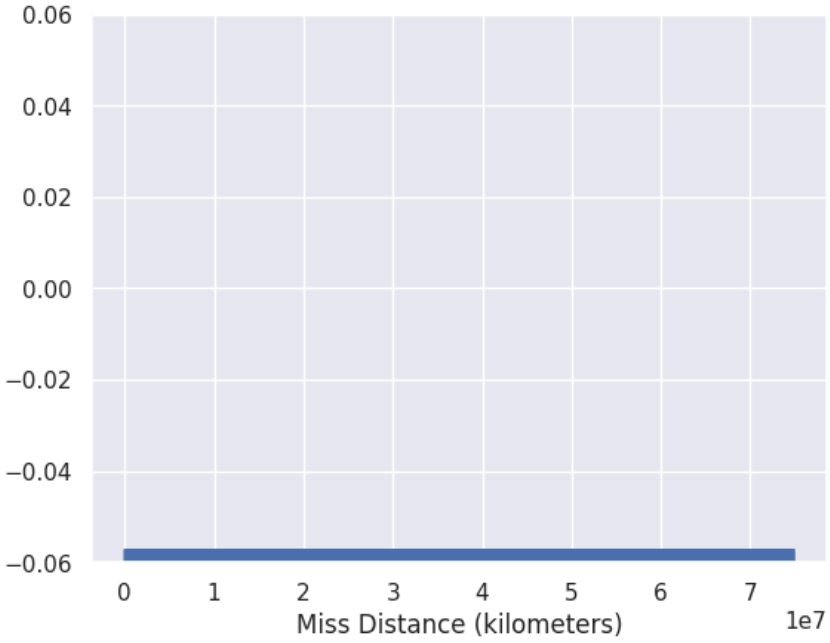
```
5879 Almeria (1992 CH1) 138911 (2001 AE2) 187026 (2005 EK70) 380818 (2005 YV128) 518810 (2010 CF19) (2008 EA8) (2012 D
```

```
stopwords = set(STOPWORDS)
```

```
wordcloud = WordCloud(background_color="White").generate(text)
plt.imshow(wordcloud, interpolation= 'bilinear')
plt.axis("off")
plt.margins(x=0, y=0)
plt.show()
```

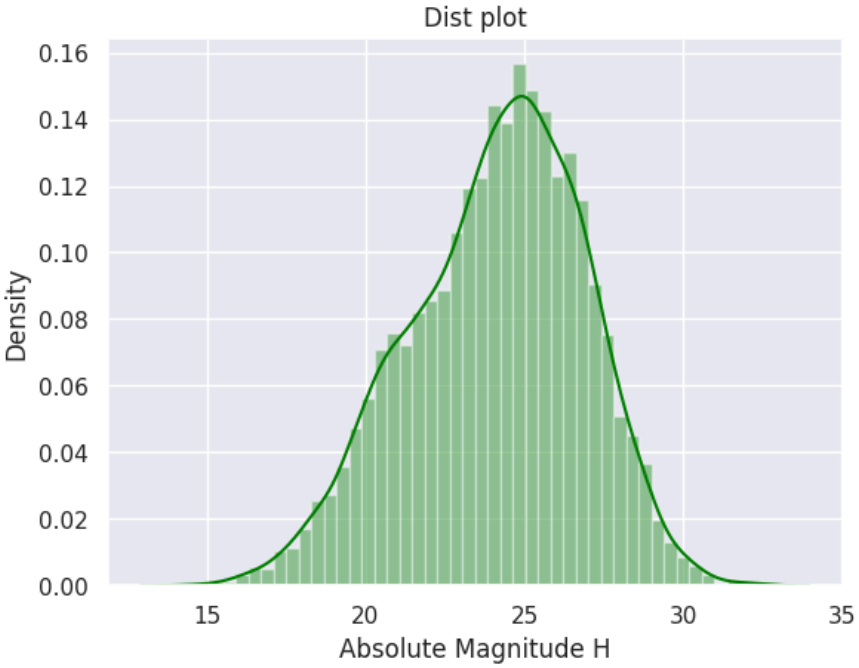
```
sns.rugplot(df['Miss Distance (kilometers)'])
```

```
<Axes: xlabel='Miss Distance (kilometers)'>
```

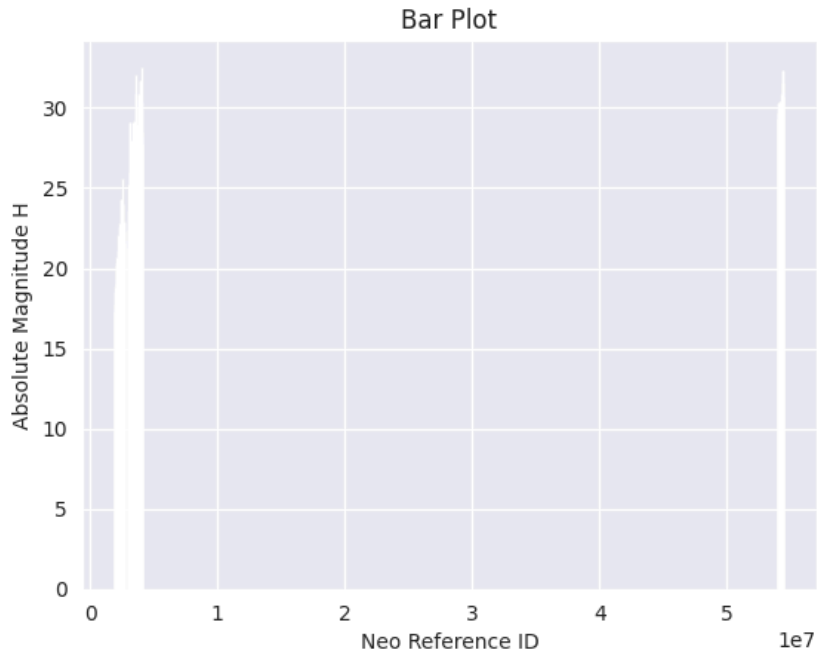



```
sns.distplot(df['Absolute Magnitude H'], kde=True, color='green').set_title('Dist plot')
```

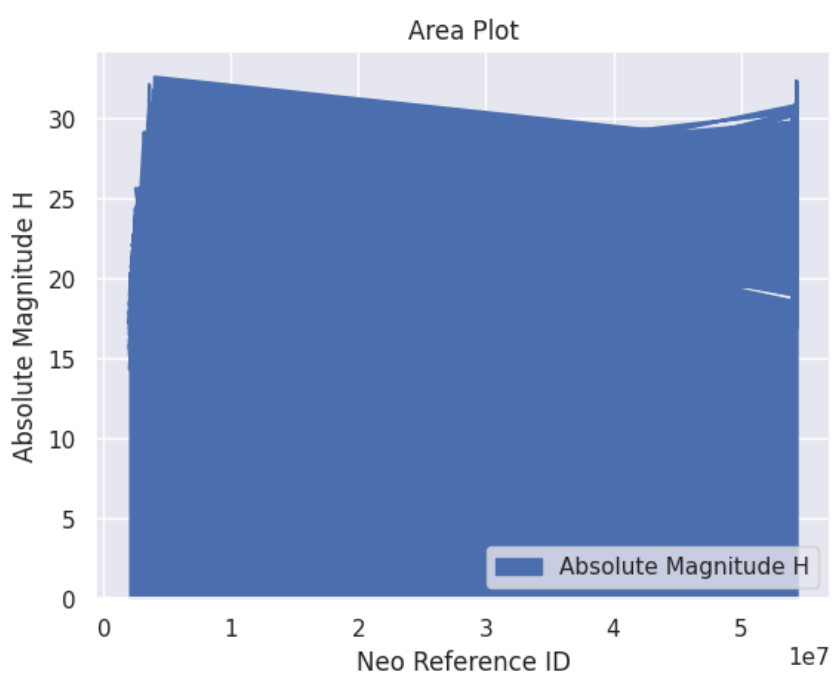
```
<ipython-input-12-4715fa789a36>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['Absolute Magnitude H'], kde=True, color='green').set_title('Dist plot')
Text(0.5, 1.0, 'Dist plot')
```




```
# prompt: create all possible types of plots using df

# **BAR PLOT**
plt.bar(df['Neo Reference ID'], df['Absolute Magnitude H'],color='blue')
plt.title('Bar Plot')
plt.xlabel('Neo Reference ID')
plt.ylabel('Absolute Magnitude H')
plt.show()
```




```
# **AREA PLOT**
df.plot.area(x='Neo Reference ID', y='Absolute Magnitude H')
plt.title('Area Plot')
plt.xlabel('Neo Reference ID')
plt.ylabel('Absolute Magnitude H')
plt.show()
```




```
sns.set_style('darkgrid')
fig, ax= plt.subplots(nrows=4, ncols=2)
fig.set_size_inches(18.5, 10.5)

sns.barplot(x='Is Potentially Hazardous Asteroid', y='Relative Velocity (km/h)', data = df,palette='plasma',estimator=np.st

sns.countplot(x='Orbiting Body', data = df,ax=ax[0,1]).set_title('Count Plot')

sns.boxplot(x='Is Potentially Hazardous Asteroid',y='Miss Distance (astronomical units)',data=df, ax=ax[1,0]).set_title('Bo

sns.violinplot(x='Is Potentially Hazardous Asteroid',y='Miss Distance (kilometers)',data=df, split=True, ax=ax[1,1]).set_ti
```

```
sns.stripplot(x='Is Potentially Hazardous Asteroid',y='Absolute Magnitude H',data=df, jitter= True,dodge=True, ax=ax[2,0]).


sns.violinplot(x='Is Potentially Hazardous Asteroid',y='Miss Distance (astronomical units)',data=df,ax=ax[2,1])

sns.boxenplot(x='Is Potentially Hazardous Asteroid',y='Absolute Magnitude H',color="b",scale="linear",data=df,ax=ax[3,0])

sns.pointplot(x='Is Potentially Hazardous Asteroid',y='Absolute Magnitude H',color="b",data=df,ax=ax[3,1])
```

```
<ipython-input-7-ebe046df78bb>:5: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.

  sns.barplot(x='Is Potentially Hazardous Asteroid', y='Relative Velocity (km/h)', da
<ipython-input-7-ebe046df78bb>:18: FutureWarning:

The `scale` parameter has been renamed to `width_method` and will be removed in v0.15
  sns.boxenplot(x='Is Potentially Hazardous Asteroid',y='Absolute Magnitude H',color=
<Axes: xlabel='Is Potentially Hazardous Asteroid', ylabel='Absolute Magnitude H'>
```