```
import pandas as pd
import numpy as np
import matplotlib.pylab as plt
import seaborn as sns
plt.style.use('ggplot')
pd.set_option('max_columns', 200)
```

```
df = pd.read_csv('coaster.csv')
```

```
df.head(3)
```

	coaster_name	Length	Speed	Location	Status	Opening date	Type	Manufacturer
0	Switchback Railway	600 ft (180 m)	6 mph (9.7 km/h)	Coney Island	Removed	June 16, 1884	Wood	LaMarcus Adna Thompson
1	Flip Flap Railway	NaN	NaN	Sea Lion Park	Removed	1895	Wood	Lina Beecher
2	Switchback Railway (Euclid Beach Park)	NaN	NaN	Cleveland, Ohio, United States	Closed	NaN	Other	NaN

3 rows × 56 columns

```
df.tail(3)
```

	coaster_name	Length	Speed	Location	Status	Opening date	Type	M
1084	Tron Lightcycle Power Run	3,169.3 ft (966.0 m)	59.3[1] mph (95.4 km/h)	Other	NaN	June 16, 2016	Steel – Launched	
1085	Tumbili	770 ft (230 m)	34 mph (55 km/h)	Kings Dominion	Under construction	NaN	Steel – 4th Dimension – Wing Coaster	
1086	Wonder Woman Flight of Courage	3,300 ft (1,000 m)	58 mph (93 km/h)	Six Flags Magic Mountain	Under construction	2022	Steel – Single-rail	

3 rows × 56 columns

```
df.shape
```

```
(1087, 56)
```

```
df.describe()
```

	Inversions	year_introduced	latitude	longitude	speed1_value	speed
count	932.000000	1087.000000	812.000000	812.000000	937.000000	937.00

```
df.columns
```

```
Index(['coaster_name', 'Length', 'Speed', 'Location', 'Status', 'Opening date',
       'Type', 'Manufacturer', 'Height restriction', 'Model', 'Height',
       'Inversions', 'Lift/launch system', 'Cost', 'Trains', 'Park section',
       'Duration', 'Capacity', 'G-force', 'Designer', 'Max vertical angle',
       'Drop', 'Soft opening date', 'Fast Lane available', 'Replaced',
       'Track layout', 'Fastrack available', 'Soft opening date.1',
       'Closing date', 'Opened', 'Replaced by', 'Website',
       'Flash Pass Available', 'Must transfer from wheelchair', 'Theme',
       'Single rider line available', 'Restraint Style',
       'Flash Pass available', 'Acceleration', 'Restraints', 'Name',
       'year_introduced', 'latitude', 'longitude', 'Type_Main',
       'opening_date_clean', 'speed1', 'speed2', 'speed1_value', 'speed1_unit',
       'speed_mph', 'height_value', 'height_unit', 'height_ft',
       'Inversions_clean', 'Gforce_clean'],
      dtype='object')
```

```
df.dtypes
```

```
coaster_name                     object
Length                           object
Speed                            object
Location                         object
Status                           object
Opening date                     object
Type                             object
Manufacturer                     object
Height restriction               object
Model                            object
Height                           object
Inversions                      float64
Lift/launch system               object
Cost                             object
Trains                           object
Park section                     object
Duration                         object
Capacity                         object
G-force                          object
Designer                         object
Max vertical angle               object
Drop                             object
Soft opening date                object
Fast Lane available              object
Replaced                         object
Track layout                     object
Fastrack available               object
Soft opening date.1              object
Closing date                     object
Opened                           object
Replaced by                      object
Website                          object
Flash Pass Available             object
Must transfer from wheelchair    object
Theme                            object
Single rider line available      object
Restraint Style                  object
Flash Pass available             object
Acceleration                     object
Restraints                       object
Name                             object
year_introduced                   int64
latitude                        float64
longitude                       float64
Type_Main                        object
opening_date_clean               object
speed1                           object
speed2                           object
speed1_value                    float64
speed1_unit                      object
speed_mph                       float64
height_value                    float64
height_unit                      object
height_ft                       float64
Inversions_clean                  int64
Gforce_clean                    float64
dtype: object
```

```
df[['coaster_name','Manufacturer', 'Opened', 'year_introduced', 'latitude', 'longitude', 'Type_Main', 'opening_date_clean', 'speed_mph'
```

	coaster_name	Manufacturer	Opened	year_introduced	latitude	longitude
0	Switchback Railway	LaMarcus Adna Thompson	NaN	1884	40.5740	-73.9780
1	Flip Flap Railway	Lina Beecher	NaN	1895	40.5780	-73.9790
2	Switchback Railway (Euclid Beach Park)	NaN	1895	1896	41.5800	-81.5700
3	Loop the Loop (Coney Island)	Edwin Prescott	NaN	1901	40.5745	-73.9780
4	Loop the Loop (Young's Pier)	Edwin Prescott	NaN	1901	39.3538	-74.4342
...	...	...	...	...	...	...
1082	American Dreier Looping	Anton Schwarzkopf	NaN	2022	NaN	NaN
1083	Pantheon (roller coaster)	Intamin	NaN	2022	37.2339	-76.6426
1084	Tron Lightcycle Power Run	Vekoma	NaN	2022	NaN	NaN

```
df.drop(['Opening date'], axis = 1)
```

	coaster_name	Length	Speed	Location	Status	Type	Manufac
1	Railway	NaN	NaN	Park	Removed	Wood	Lina Be

```
df.shape
```

```
(1087, 56)
```

```
df.dtypes
```

```
coaster_name                     object
Length                           object
Speed                            object
Location                         object
Status                           object
Opening date                     object
Type                             object
Manufacturer                     object
Height restriction               object
Model                            object
Height                           object
Inversions                      float64
Lift/launch system               object
Cost                             object
Trains                           object
Park section                     object
Duration                         object
Capacity                         object
G-force                          object
Designer                         object
Max vertical angle               object
Drop                             object
Soft opening date                object
Fast Lane available              object
Replaced                         object
Track layout                     object
Fastrack available               object
Soft opening date.1              object
Closing date                     object
Opened                           object
Replaced by                      object
Website                          object
Flash Pass Available             object
Must transfer from wheelchair    object
Theme                            object
Single rider line available      object
Restraint Style                  object
Flash Pass available             object
Acceleration                     object
Restraints                       object
Name                             object
year_introduced                   int64
latitude                        float64
longitude                       float64
Type_Main                        object
opening_date_clean               object
speed1                           object
speed2                           object
speed1_value                    float64
speed1_unit                      object
speed_mph                       float64
height_value                    float64
height_unit                      object
height_ft                       float64
Inversions_clean                  int64
Gforce_clean                    float64
dtype: object
```

```
df['opening_date_clean'] = pd.to_datetime(df['opening_date_clean'])
```

```
df.columns.nunique()
```

```
56
```

```
df['opening_date_clean'].nunique()
```

```
602
```

```
df = df.rename(columns = {'coaster_name' : 'Coaster_Name',
                      'year_introduced' : 'Year_Introduced',
                      'opening_date_clean' : 'Opening_Date',
                      'speed_mph' : 'Speed_mph',
                      'height_ft' : 'Height_ft',
                      'Inversions_clean' : 'Inversions',
                      'Gforce_clean' : 'Gforce'})
```

```
df.head()
```

	Coaster_Name	Length	Speed	Location	Status	Opening date	Type	Manufact
0	Switchback Railway	600 ft (180 m)	6 mph (9.7 km/h)	Coney Island	Removed	June 16, 1884	Wood	LaM Thom
1	Flip Flap Railway	NaN	NaN	Sea Lion Park	Removed	1895	Wood	Lina Be
2	Switchback Railway (Euclid Beach Park)	NaN	NaN	Cleveland, Ohio, United States	Closed	NaN	Other	
3	Loop the Loop (Coney Island)	NaN	NaN	Other	Removed	1901	Steel	Edwin Pre
4	Loop the Loop (Young's Pier)	NaN	NaN	Other	Removed	1901	Steel	Edwin Pre

5 rows × 56 columns

```
df.isna().sum()
```

```
Coaster_Name                    0
Length                        134
Speed                         150
Location                        0
Status                        213
Opening date                  250
Type                            0
Manufacturer                   59
Height restriction            256
Model                         343
Height                        122
Inversions                    155
Lift/launch system            292
Cost                          705
Trains                        369
Park section                  600
Duration                      322
Capacity                      512
G-force                       725
Designer                      509
Max vertical angle            730
Drop                          593
Soft opening date             991
Fast Lane available          1018
Replaced                      914
Track layout                  752
Fastrack available           1068
Soft opening date.1           991
Closing date                  851
Opened                       1060
Replaced by                   999
Website                      1000
Flash Pass Available         1037
Must transfer from wheelchair 981
Theme                        1043
Single rider line available  1006
Restraint Style              1065
Flash Pass available         1041
Acceleration                 1027
Restraints                   1063
Name                         1052
Year_Introduced                 0
latitude                      275
```

```
longitude          275
Type_Main            0
Opening_Date       250
speed1             150
speed2             152
speed1_value       150
speed1_unit        150
Speed_mph          150
height_value       122
height_unit        122
Height_ft          916
Inversions           0
Gforce             725
dtype: int64
```

```
df.loc[~df.duplicated(subset = ['Coaster_Name', 'Location', 'Opening_Date'])]\
.reset_index(drop = True).copy()
```

	Coaster_Name	Length	Speed	Location	Status	Opening date	
0	Switchback Railway	600 ft (180 m)	6 mph (9.7 km/h)	Coney Island	Removed	June 16, 1884	V
1	Flip Flap Railway	NaN	NaN	Sea Lion Park	Removed	1895	V
2	Switchback Railway (Euclid Beach Park)	NaN	NaN	Cleveland, Ohio, United States	Closed	NaN	(
3	Loop the Loop (Coney Island)	NaN	NaN	Other	Removed	1901	
4	Loop the Loop (Young's Pier)	NaN	NaN	Other	Removed	1901	
...	...	...	...	...	...	...	
985	Ice Breaker (roller coaster)	1,900 ft (580 m)	52 mph (84 km/h)	SeaWorld Orlando	Under construction	February 2022	St Laun
986	Leviathan (Sea World)	1,000.0 m (3,280.8 ft)	80.0 km/h (49.7 mph)	Sea World	Under construction	Easter 2022	V
987	Pantheon (roller coaster)	3,328 ft (1,014 m)	73 mph (117 km/h)	Busch Gardens Williamsburg	Under construction	2022	St Laun
988	Tumbili	770 ft (230 m)	34 mph (55 km/h)	Kings Dominion	Under construction	NaN	Steel Dimens Co
989	Wonder Woman Flight of Courage	3,300 ft (1,000 m)	58 mph (93 km/h)	Six Flags Magic Mountain	Under construction	2022	St Singl

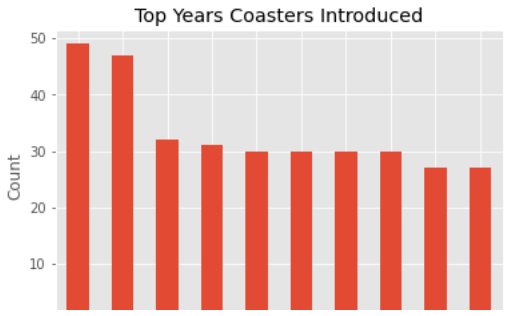
990 rows × 56 columns

```
df.shape
```

```
(1087, 56)
```

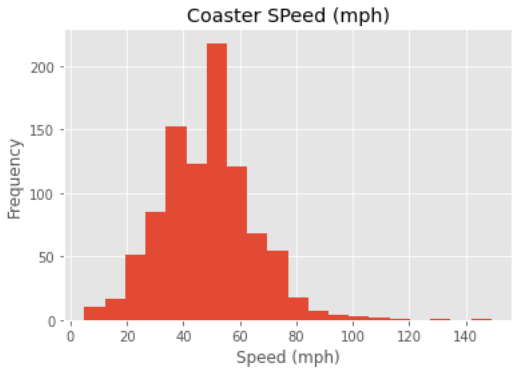
```
ax = df['Year_Introduced'].value_counts().head(10).plot(kind = 'bar', title = 'Top Years Coasters Introduced')
ax.set_xlabel('Year Introduced')
ax.set_ylabel('Count')
```

```
Text(0, 0.5, 'Count')
```



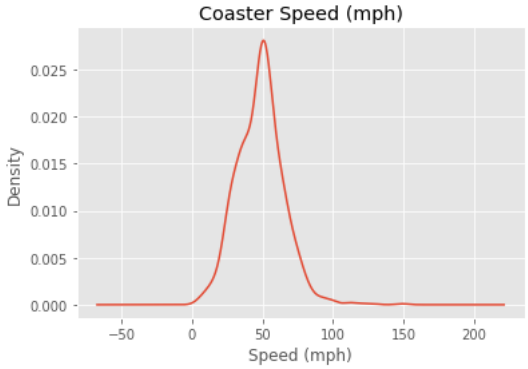

```
ax = df['Speed_mph'].plot(kind = 'hist', bins = 20, title = 'Coaster SPeed (mph)')
ax.set_xlabel('Speed (mph)')
```

```
Text(0.5, 0, 'Speed (mph)')
```

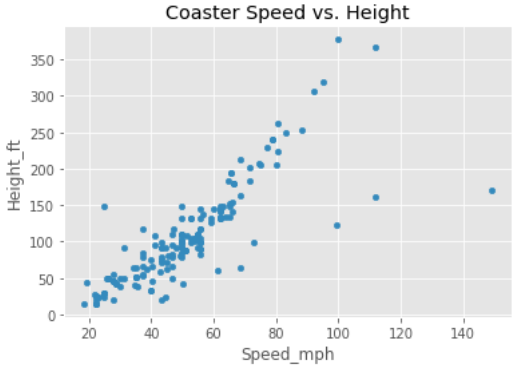



```
ax = df['Speed_mph'].plot(kind = 'kde', title = 'Coaster Speed (mph)')
ax.set_xlabel('Speed (mph)')
```

```
Text(0.5, 0, 'Speed (mph)')
```

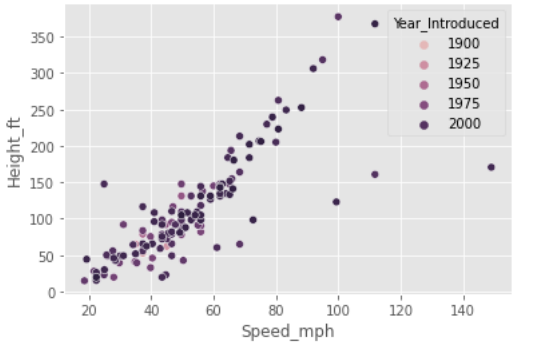



```
df.plot(kind = 'scatter', x = 'Speed_mph', y = 'Height_ft', title = 'Coaster Speed vs. Height')
plt.show()
```




```
sns.scatterplot(x = 'Speed_mph', y = 'Height_ft', hue = 'Year_Introduced', data = df)
```

```
<AxesSubplot:xlabel='Speed_mph', ylabel='Height_ft'>
```



✓ 0s completed at 12:55 PM