

In [ ]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
!pip install nbconvert
```

```
Requirement already satisfied: nbconvert in c:\users\harish vasanth\anaconda3\lib\site-packages (6.1.0)
Requirement already satisfied: pandocfilters>=1.4.1 in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbconvert) (1.4.3)
Requirement already satisfied: jupyterlab-pygments in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbconvert) (0.1.2)
Requirement already satisfied: defusedxml in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbconvert) (0.7.1)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbconvert) (0.5.3)
Requirement already satisfied: bleach in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbconvert) (4.0.0)
Requirement already satisfied: jupyter-core in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbconvert) (4.8.1)
Requirement already satisfied: pygments>=2.4.1 in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbconvert) (2.10.0)
Requirement already satisfied: entrypoints>=0.2.2 in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbconvert) (0.3)
Requirement already satisfied: traitlets>=5.0 in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbconvert) (5.1.0)
Requirement already satisfied: mistune<2,>=0.8.1 in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbconvert) (0.8.4)
Requirement already satisfied: testpath in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbconvert) (0.5.0)
Requirement already satisfied: jinja2>=2.4 in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbconvert) (2.11.3)
Requirement already satisfied: nbformat>=4.4 in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbconvert) (5.1.3)
Requirement already satisfied: MarkupSafe>=0.23 in c:\users\harish vasanth\anaconda3\lib\site-packages (from jinja2>=2.4->nbconvert) (1.1.1)
Requirement already satisfied: async-generator in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbclient<0.6.0,>=0.5.0->nbconvert) (1.10)
Requirement already satisfied: jupyter-client>=6.1.5 in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbclient<0.6.0,>=0.5.0->nbconvert) (6.1.12)
Requirement already satisfied: nest-asyncio in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbclient<0.6.0,>=0.5.0->nbconvert) (1.5.1)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\harish vasanth\anaconda3\lib\site-packages (from jupyter-client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconvert) (2.8.2)
Requirement already satisfied: pyzmq>=13 in c:\users\harish vasanth\anaconda3\lib\site-packages (from jupyter-client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconvert) (22.2.1)
Requirement already satisfied: tornado>=4.1 in c:\users\harish vasanth\anaconda3\lib\site-packages (from jupyter-client>=6.1.5->nbclient<0.6.0,>=0.5.0->nbconvert) (6.1)
Requirement already satisfied: pywin32>=1.0 in c:\users\harish vasanth\anaconda3\lib\site-packages (from jupyter-core->nbconvert) (228)
Requirement already satisfied: ipython-genutils in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbformat>=4.4->nbconvert) (0.2.0)
Requirement already satisfied: jsonschema!=2.5.0,>=2.4 in c:\users\harish vasanth\anaconda3\lib\site-packages (from nbformat>=4.4->nbconvert) (3.2.0)
Requirement already satisfied: attrs>=17.4.0 in c:\users\harish vasanth\anaconda3\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (21.2.0)
Requirement already satisfied: six>=1.11.0 in c:\users\harish vasanth\anaconda3\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (1.16.0)
Requirement already satisfied: pyrsistent>=0.14.0 in c:\users\harish vasanth\anaconda3\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (0.18.0)
```

Requirement already satisfied: setuptools in c:\users\harish vasanth\anaconda3\lib\site-packages (from jsonschema!=2.5.0,>=2.4->nbformat>=4.4->nbconvert) (67.6.0)  
 Requirement already satisfied: webencodings in c:\users\harish vasanth\anaconda3\lib\site-packages (from bleach->nbconvert) (0.5.1)  
 Requirement already satisfied: packaging in c:\users\harish vasanth\anaconda3\lib\site-packages (from bleach->nbconvert) (21.0)  
 Requirement already satisfied: pyparsing>=2.0.2 in c:\users\harish vasanth\anaconda3\lib\site-packages (from packaging->bleach->nbconvert) (3.0.4)

In [ ]:

In [ ]:

df = pd.read\_csv('coaster\_db.csv')

## Exploratory analysis

1 . Distinguish Numerical and categorical features.

2 . Handle missing and Null values

3 . Identify / remove outliers in the data

4 . Further clean the data.

In [ ]:

df.head(10)

Out[ ]:

	coaster_name	Length	Speed	Location	Status	Opening date	Type	Manufacturer	Height restrict
0	Switchback Railway	600 ft (180 m)	6 mph (9.7 km/h)	Coney Island	Removed	June 16, 1884	Wood	LaMarcus Adna Thompson	NaN
1	Flip Flap Railway	NaN	NaN	Sea Lion Park	Removed	1895	Wood	Lina Beecher	NaN
2	Switchback Railway (Euclid Beach Park)	NaN	NaN	Cleveland, Ohio, United States	Closed	NaN	Other	NaN	NaN
3	Loop the Loop (Coney Island)	NaN	NaN	Other	Removed	1901	Steel	Edwin Prescott	NaN
4	Loop the Loop (Young's Pier)	NaN	NaN	Other	Removed	1901	Steel	Edwin Prescott	NaN
5	Cannon Coaster	NaN	NaN	Coney Island	Removed	1902	Wood	George Francis Meyer	NaN
6	Leap-The-Dips	1,452 ft (443 m)	10 mph (16 km/h)	Lakemont Park	Operating	1902	Wood – Side friction	Federal Construction Company	NaN
7	Figure Eight (Euclid Beach Park)	NaN	NaN	Cleveland, Ohio, United States	Closed	NaN	Other	NaN	NaN

	coaster_name	Length	Speed	Location	Status	Opening date	Type	Manufacturer	Height restriction
8	Drop the Dip	NaN	NaN	Coney Island	Removed	June 6, 1907	Other	Arthur Jarvis	N
9	Scenic Railway (Euclid Beach Park)	NaN	NaN	Cleveland, Ohio, United States	Closed	NaN	Other	NaN	N

10 rows × 56 columns

In [ ]: df.describe()

	Inversions	year_introduced	latitude	longitude	speed1_value	speed_mph	height_value
<b>count</b>	932.000000	1087.000000	812.000000	812.000000	937.000000	937.000000	965.000000
<b>mean</b>	1.547210	1994.986201	38.373484	-41.595373	53.850374	48.617289	89.575171
<b>std</b>	2.114073	23.475248	15.516596	72.285227	23.385518	16.678031	136.246444
<b>min</b>	0.000000	1884.000000	-48.261700	-123.035700	5.000000	5.000000	4.000000
<b>25%</b>	0.000000	1989.000000	35.031050	-84.552200	40.000000	37.300000	44.000000
<b>50%</b>	0.000000	2000.000000	40.289800	-76.653600	50.000000	49.700000	79.000000
<b>75%</b>	3.000000	2010.000000	44.799600	2.778100	63.000000	58.000000	113.000000
<b>max</b>	14.000000	2022.000000	63.230900	153.426500	240.000000	149.100000	3937.000000

In [ ]: 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

```

Get only the numeric data first to handle.

In [ ]: `df._get_numeric_data()`

Out[ ]:

	<b>Inversions</b>	<b>year_introduced</b>	<b>latitude</b>	<b>longitude</b>	<b>speed1_value</b>	<b>speed_mph</b>	<b>height_value</b>	<b>heig</b>
<b>0</b>	NaN	1884	40.5740	-73.9780	6.0	6.0	50.0	
<b>1</b>	1.0	1895	40.5780	-73.9790	NaN	NaN	NaN	
<b>2</b>	NaN	1896	41.5800	-81.5700	NaN	NaN	NaN	
<b>3</b>	1.0	1901	40.5745	-73.9780	NaN	NaN	NaN	
<b>4</b>	1.0	1901	39.3538	-74.4342	NaN	NaN	NaN	
...	...	...	...	...	...	...	...	...
<b>1082</b>	3.0	2022	NaN	NaN	53.0	53.0	111.0	
<b>1083</b>	2.0	2022	37.2339	-76.6426	73.0	73.0	178.0	
<b>1084</b>	0.0	2022	NaN	NaN	59.3	59.3	78.1	
<b>1085</b>	0.0	2022	NaN	NaN	34.0	34.0	112.0	
<b>1086</b>	3.0	2022	NaN	NaN	58.0	58.0	131.0	

1087 rows × 10 columns



```
In [ ]: df.select_dtypes(include=['object']).columns.tolist()
```

```
Out[ ]: ['coaster_name',
 'Length',
 'Speed',
 'Location',
 'Status',
 'Opening date',
 'Type',
 'Manufacturer',
 'Height restriction',
 'Model',
 'Height',
 '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',
 'Type_Main',
 'opening_date_clean',
 'speed1',
 'speed2',
 'speed1_unit',
 'height_unit']
```

```
In [ ]: df = df[['coaster_name', 'Location', 'Status', 'Manufacturer', 'year_introduced', 'latit
```

Let's get the factors that we want to analyse separately first.

```
In [ ]: df
```

	coaster_name	Location	Status	Manufacturer	year_introduced	latitude	longitude
0	Switchback Railway	Coney Island	Removed	LaMarcus Adna Thompson	1884	40.5740	-73.9780

	coaster_name	Location	Status	Manufacturer	year_introduced	latitude	longitude
1	Flip Flap Railway	Sea Lion Park	Removed	Lina Beecher	1895	40.5780	-73.9790
2	Switchback Railway (Euclid Beach Park)	Cleveland, Ohio, United States	Closed	NaN	1896	41.5800	-81.5700
3	Loop the Loop (Coney Island)	Other	Removed	Edwin Prescott	1901	40.5745	-73.9780
4	Loop the Loop (Young's Pier)	Other	Removed	Edwin Prescott	1901	39.3538	-74.4342
...	...	...	...	...	...	...	...
1082	American Dreier Looping	Other	NaN	Anton Schwarzkopf	2022	NaN	NaN
1083	Pantheon (roller coaster)	Busch Gardens Williamsburg	Under construction	Intamin	2022	37.2339	-76.6426
1084	Tron Lightcycle Power Run	Other	NaN	Vekoma	2022	NaN	NaN
1085	Tumbili	Kings Dominion	Under construction	S&S – Sansei Technologies	2022	NaN	NaN
1086	Wonder Woman Flight of Courage	Six Flags Magic Mountain	Under construction	Rocky Mountain Construction	2022	NaN	NaN

1087 rows × 13 columns

```
In [ ]: df.isna().sum() # Find the factors with NA values.
```

```
Out[ ]: coaster_name      0
Location          0
Status         213
Manufacturer     59
year_introduced   0
latitude        275
longitude        275
Type_Main        0
opening_date_clean 250
speed_mph       150
height_ft       916
Inversions_clean  0
Gforce_clean     725
dtype: int64
```

```
In [ ]: df.loc[df.duplicated()] ## Seems as though there is no repeated rows.
## TO handle repeated rows , we could have removed : first duplicate , last duplicat
```

```
Out[ ]: coaster_name  Location  Status  Manufacturer  year_introduced  latitude  longitude  Type_Main  o
```

```
In [ ]: df.loc[df.duplicated(subset=['coaster_name'])]
```

Out[ ]:

	coaster_name	Location	Status	Manufacturer	year_introduced	latitude	longitude	T
43	Crystal Beach Cyclone	Crystal Beach Park	Removed	Traver Engineering	1927	42.8617	-79.0598	
60	Derby Racer	Revere Beach	Removed	Fred W. Pearce	1937	42.4200	-70.9860	
61	Blue Streak (Conneaut Lake)	Conneaut Lake Park	Closed		Nan	1938	41.6349	-80.3180
167	Big Thunder Mountain Railroad	Other	Nan	Arrow Development (California and Florida)Dyna...	1980		Nan	
237	Thunder Run (Canada's Wonderland)	Canada's Wonderland	Operating	Mack Rides	1986	43.8427	-79.5423	
...	...	...	...	...	...	...	...	...
1063	Lil' Devil Coaster	Six Flags Great Adventure	Operating	Zamperla	2021	40.1343	-74.4434	
1064	Little Dipper (Conneaut Lake Park)	Conneaut Lake Park	Operating	Allan Herschell Company	2021	41.6343	-80.3165	
1080	Iron Gwazi	Busch Gardens Tampa Bay	Under construction	Rocky Mountain Construction	2022	28.0339	-82.4231	
1082	American Dreier Looping	Other	Nan	Anton Schwarzkopf	2022		Nan	
1084	Tron Lightcycle Power Run	Other	Nan	Vekoma	2022		Nan	

97 rows × 13 columns

```
In [ ]: df.query('coaster_name == "Crystal Beach Cyclone"') # Looking at duplicate val
```

	coaster_name	Location	Status	Manufacturer	year_introduced	latitude	longitude	Type_Mai
39	Crystal Beach Cyclone	Crystal Beach Park	Removed	Traver Engineering	1926	42.8617	-79.0598	Woo
43	Crystal Beach Cyclone	Crystal Beach Park	Removed	Traver Engineering	1927	42.8617	-79.0598	Woo



```
In [ ]: df.drop_duplicates(subset=['coaster_name', 'Location', 'opening_date_clean'], keep='fir
```

	coaster_name	Location	Status	Manufacturer	year_introduced	latitude	longitude
0	Switchback Railway	Coney Island	Removed	LaMarcus Adna Thompson	1884	40.5740	-73.9780
1	Flip Flap Railway	Sea Lion Park	Removed	Lina Beecher	1895	40.5780	-73.9790
2	Switchback Railway (Euclid Beach Park)	Cleveland, Ohio, United States	Closed	NaN	1896	41.5800	-81.5700
3	Loop the Loop (Coney Island)	Other	Removed	Edwin Prescott	1901	40.5745	-73.9780
4	Loop the Loop (Young's Pier)	Other	Removed	Edwin Prescott	1901	39.3538	-74.4342
...	...	...	...	...	...	...	...
1079	Ice Breaker (roller coaster)	SeaWorld Orlando	Under construction	Premier Rides	2022	28.4088	-81.4633
1081	Leviathan (Sea World)	Sea World	Under construction	Martin & Vleminckx	2022	-27.9574	153.4263
1083	Pantheon (roller coaster)	Busch Gardens Williamsburg	Under construction	Intamin	2022	37.2339	-76.6426
1085	Tumbili	Kings Dominion	Under construction	S&S – Sansei Technologies	2022	NaN	NaN
1086	Wonder Woman Flight of Courage	Six Flags Magic Mountain	Under construction	Rocky Mountain Construction	2022	NaN	NaN

990 rows × 13 columns



```
In [ ]: df.query('coaster_name == "Crystal Beach Cyclone"')
```

	coaster_name	Location	Status	Manufacturer	year_introduced	latitude	longitude	Type_Mai
39	Crystal Beach Cyclone	Crystal Beach Park	Removed	Traver Engineering	1926	42.8617	-79.0598	Woo
43	Crystal Beach Cyclone	Crystal Beach Park	Removed	Traver Engineering	1927	42.8617	-79.0598	Woo



In [ ]:

## Visual Analytics

In [ ]:

```
df['year_introduced'].value_counts()
```

Out[ ]:

1999	49
2000	47
1998	32
2011	31
2001	30
..	
1956	1
1959	1
1961	1
1895	1
1884	1

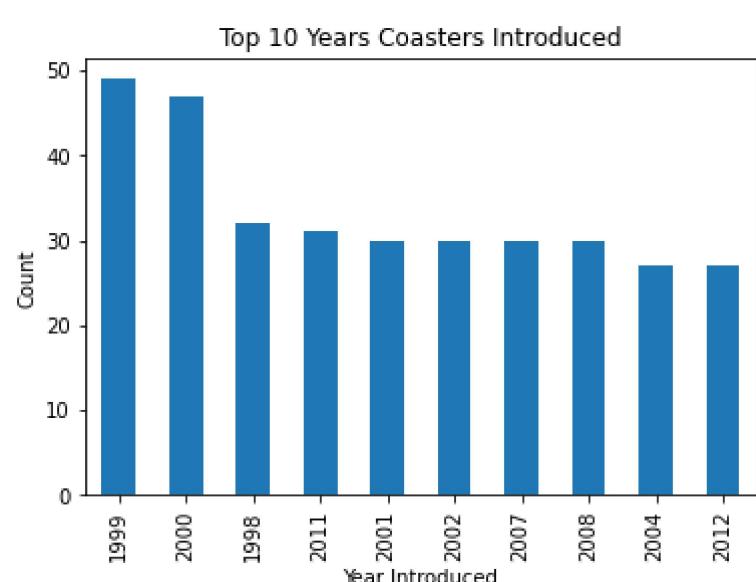
Name: year\_introduced, Length: 101, dtype: int64

In [ ]:

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

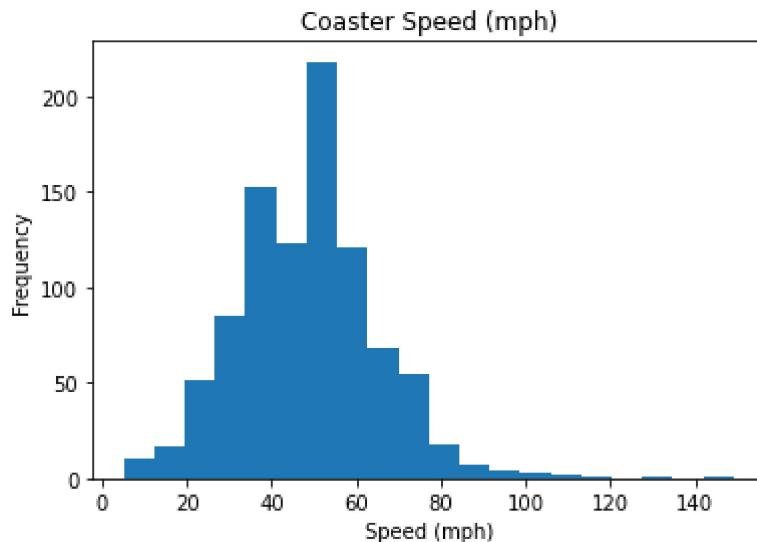
Out[ ]:

Text(0, 0.5, 'Count')



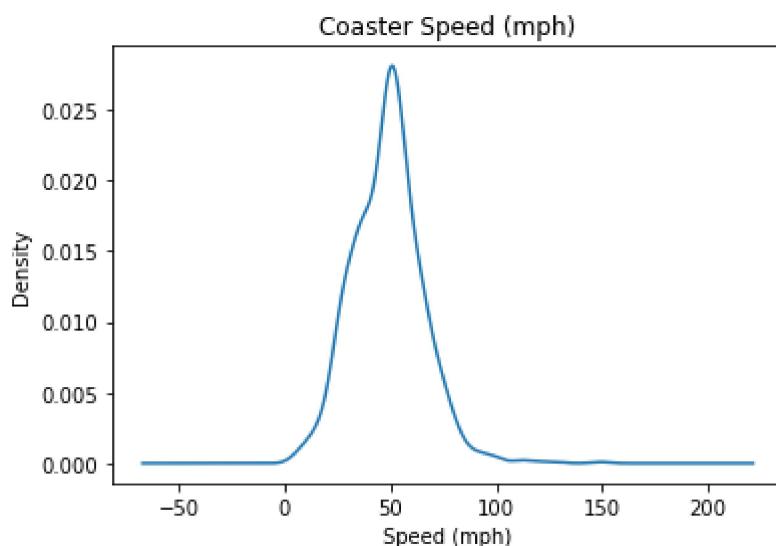
```
In [ ]: ax = df['speed_mph'].plot(kind='hist',
                                bins=20,
                                title='Coaster Speed (mph)')
ax.set_xlabel('Speed (mph)')
```

Out[ ]: Text(0.5, 0, 'Speed (mph)')



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

Out[ ]: Text(0.5, 0, 'Speed (mph)')



Since coaster speed is highly concentrated around the center with a bell curve distribution, we can replace the rows with null values with the average coaster speed.

```
In [ ]: df['speed_mph'].fillna((df['speed_mph'].mean()), inplace=True)
```

```
In [ ]: df
```

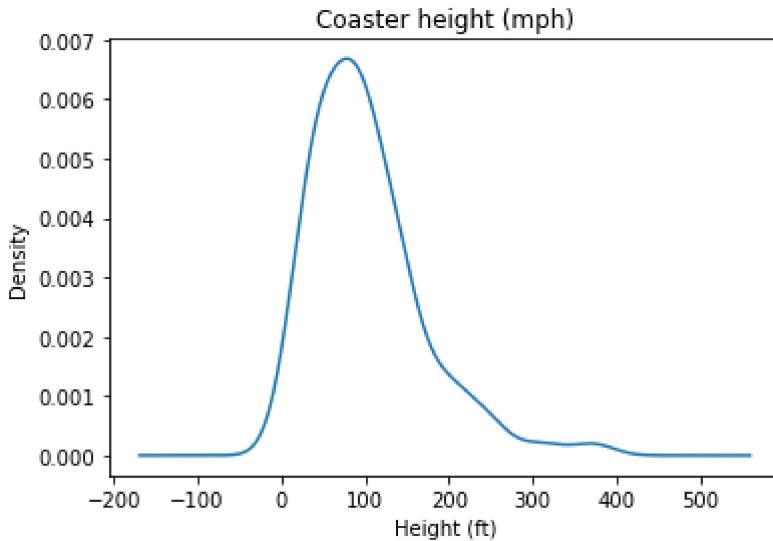
Out[ ]:	coaster_name	Location	Status	Manufacturer	year_introduced	latitude	longitude
0	Switchback Railway	Coney Island	Removed	LaMarcus Adna Thompson	1884	40.5740	-73.9780
1	Flip Flap Railway	Sea Lion Park	Removed	Lina Beecher	1895	40.5780	-73.9790
2	Switchback Railway (Euclid Beach Park)	Cleveland, Ohio, United States	Closed	NaN	1896	41.5800	-81.5700
3	Loop the Loop (Coney Island)	Other	Removed	Edwin Prescott	1901	40.5745	-73.9780
4	Loop the Loop (Young's Pier)	Other	Removed	Edwin Prescott	1901	39.3538	-74.4342
...	...	...	...	...	...	...	...
1082	American Dreier Looping	Other	NaN	Anton Schwarzkopf	2022	NaN	NaN
1083	Pantheon (roller coaster)	Busch Gardens Williamsburg	Under construction	Intamin	2022	37.2339	-76.6426
1084	Tron Lightcycle Power Run	Other	NaN	Vekoma	2022	NaN	NaN
1085	Tumbili	Kings Dominion	Under construction	S&S – Sansei Technologies	2022	NaN	NaN
1086	Wonder Woman Flight of Courage	Six Flags Magic Mountain	Under construction	Rocky Mountain Construction	2022	NaN	NaN

1087 rows × 13 columns



```
In [ ]:
ax = df['height_ft'].plot(kind='kde',
                           title='Coaster height (mph)')
ax.set_xlabel('Height (ft)')
```

```
Out[ ]: Text(0.5, 0, 'Height (ft)')
```



**As coaster height is a bit more skewed , let us just remove the na valued rows.**

In [ ]:

In [ ]:

```
df = df.dropna(subset=['height_ft'])
df
```

Out[ ]:

	coaster_name	Location	Status	Manufacturer	year_introduced	latitude	longitude
15	The Great Scenic Railway	Luna Park Melbourne	Operating	LaMarcus Thompson	1912	NaN	NaN
25	Roller Coaster (Lagoon)	Lagoon Amusement Park	Operating	NaN	1921	40.9849	-111.8951
51	Dragon Coaster (Playland)	Playland	Operating	NaN	1929	40.9675	-73.6739
53	Yankee Cannonball	Canobie Lake Park	Operating	Philadelphia Toboggan Coasters	1930	42.7953	-71.2492
71	Big Dipper (Battersea Park)	Battersea Park	Removed	NaN	1951	51.4794	-0.1569
...	...	...	...	...	...	...	...
1046	Zadra (roller coaster)	Energylandia	Operating	Rocky Mountain Construction	2019	50.0019	19.4031
1049	Max & Moritz (roller coaster)	Efteling	Operating	Mack Rides	2020	51.6481	5.0474
1052	Pitts Special (roller coaster)	PowerPark	Operating	Gerstlauer	2020	63.2307	22.8545
1065	Looping Star (Clacton Pier)	Clacton Pier	Operating	Pinfari	2021	51.7864	-1.1550

	coaster_name	Location	Status	Manufacturer	year_introduced	latitude	longitude
1081	Leviathan (Sea World)	Sea World	Under construction	Martin & Vleminckx	2022	-27.9574	153.4263

171 rows × 13 columns

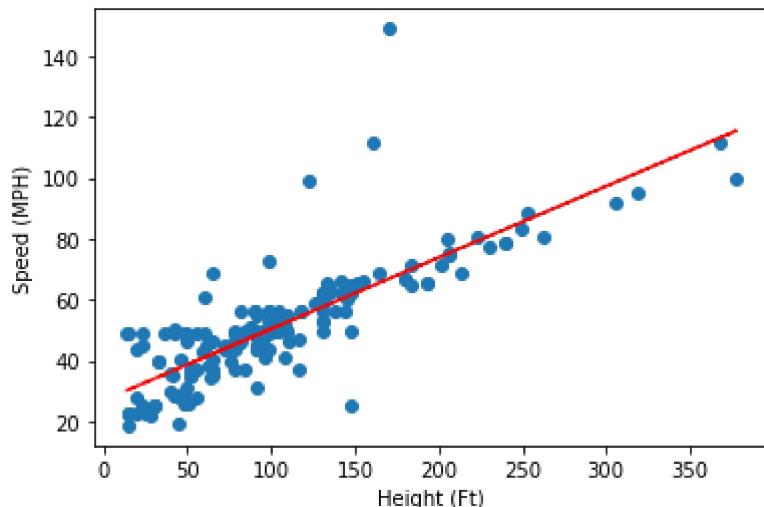
## Analyse speed vs height

In [ ]:

```
y = df['speed_mph']
x = df['height_ft']
z = np.polyfit(x, y, 1)
p = np.poly1d(z)
plt.plot(x,p(x), "r--")

plt.scatter(x, y)
plt.ylabel("Speed (MPH)")
plt.xlabel("Height (Ft)")

plt.show()
```



In [ ]:

```
# as the percentage of Gforce rows being NaN is very high, we will just remove them,
df = df.dropna(subset=['Gforce_clean'])
```

In [ ]:

df

Out[ ]:

	coaster_name	Location	Status	Manufacturer	year_introduced	latitude	longitude
183	Python (Efteling)	Efteling	Operating	Vekoma	1981	51.6469	5.0536
225	Mindbender (Galaxyland)	Galaxyland	closed for maintenance as of july 30 no reopen...	Anton Schwarzkopf	1985	53.5236	-113.6203
247	Lisebergbanan	Liseberg	Operating	Zierer	1987	57.6950	11.9925
272	Eurosat - CanCan Coaster	Europa-Park	Operating	Mack Rides	1989	NaN	NaN

	coaster_name	Location	Status	Manufacturer	year_introduced	latitude	longitude
275	Goudurix	Parc Astérix	Operating	Vekoma	1989	49.1314	2.5714
...	...	...	...	...	...	...	...
982	DC Rivals HyperCoaster	Warner Bros. Movie World	Operating	Mack Rides	2017	-27.9080	153.3140
1004	Hyperion (roller coaster)	Energylandia	Operating	Intamin	2018	NaN	NaN
1031	Hakugei (roller coaster)	Nagashima Spa Land	Operating	Rocky Mountain Construction	2019	35.0303	136.7347
1046	Zadra (roller coaster)	Energylandia	Operating	Rocky Mountain Construction	2019	50.0019	19.4031
1052	Pitts Special (roller coaster)	PowerPark	Operating	Gerstlauer	2020	63.2307	22.8545

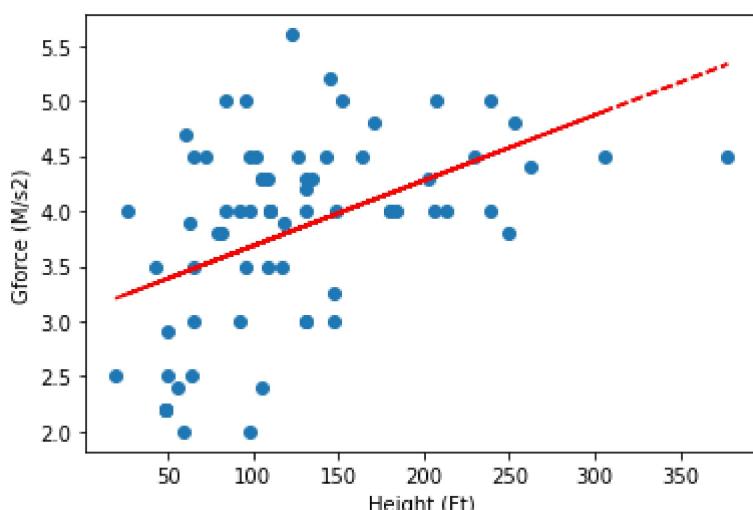
71 rows × 13 columns

### Analyse Gforce vs height

```
In [ ]:
y = df['Gforce_clean']
x = df['height_ft']
z = np.polyfit(x, y, 1)
p = np.poly1d(z)
plt.plot(x,p(x),"r--")

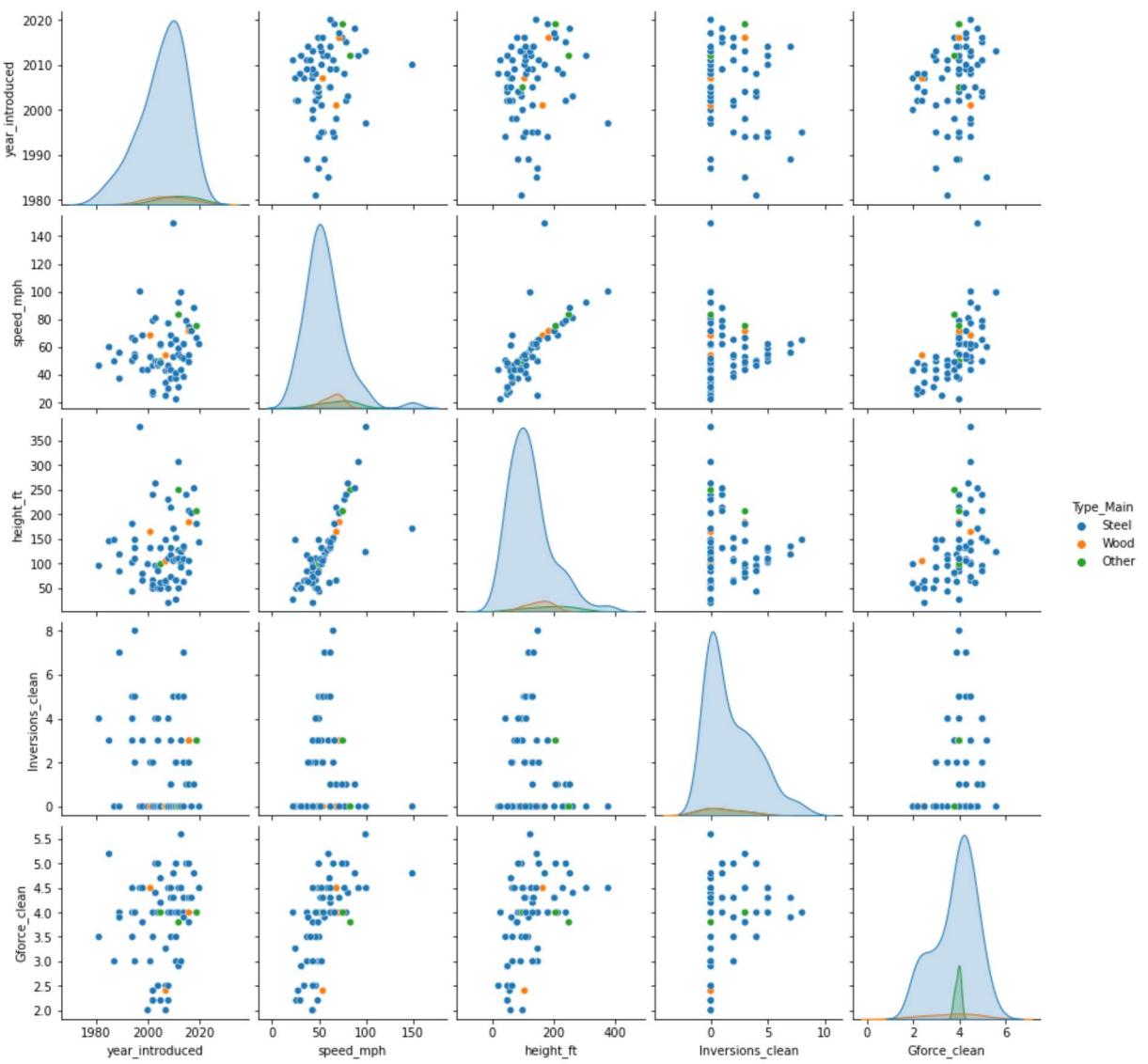
plt.scatter(x, y)
plt.ylabel("Gforce (M/s2)")
plt.xlabel("Height (Ft)")

plt.show()
```



```
In [ ]:
sns.pairplot(df,
             vars=['year_introduced', 'speed_mph',
                   'height_ft', 'Inversions_clean', 'Gforce_clean'],
             hue='Type_Main')
plt.show()
```

## Rollercoaster



In [ ]:

In [ ]: