

In []:

```
Name: Meet Hiteshkumar Trivedi
Student_id; N01520331
```

In [4]:

```
# importing libraries
import numpy as np
import pandas as pd
import matplotlib as plt
```

In [21]:

```
import seaborn as sns

%matplotlib inline
```

In [12]:

```
# importing dataset
wine = pd.read_csv('E:\Programming\Humber college\Humber Sem 2\Data Analytics\Week-5\Assign
```

In [13]:

```
wine.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1143 entries, 0 to 1142
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
---  -
0   fixed acidity          1142 non-null   float64
1   volatile acidity       1139 non-null   float64
2   citric acid            1140 non-null   float64
3   residual sugar         1143 non-null   float64
4   chlorides              1143 non-null   float64
5   free sulfur dioxide    1143 non-null   float64
6   total sulfur dioxide   1143 non-null   float64
7   density                1141 non-null   float64
8   pH                    1143 non-null   float64
9   sulphates              1143 non-null   float64
10  alcohol                1143 non-null   float64
11  quality                1141 non-null   float64
12  Id                     1143 non-null   int64
dtypes: float64(12), int64(1)
memory usage: 116.2 KB
```

In [10]:

```
wine.head()
```

Out[10]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcoh
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9

In [17]:

```
result_sorted = wine.sort_values(by = ["alcohol"], inplace =False, ascending=False)
result_sorted.head()
```

Out[17]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alc
462	15.9	0.36	0.65	7.5	0.096	22.0	71.0	0.99760	2.98	0.84	
329	8.8	0.46	0.45	2.6	0.065	7.0	18.0	0.99470	3.32	0.79	
98	5.2	0.34	0.00	1.8	0.050	27.0	63.0	0.99160	3.68	0.79	
898	5.0	0.38	0.01	1.6	0.048	26.0	60.0	0.99084	3.70	0.75	
419	5.0	0.42	0.24	2.0	0.060	19.0	50.0	0.99170	3.72	0.74	

In [16]:

```
density = wine['density']
density.dtypes
```

Out[16]:

```
dtype('float64')
```

In [19]:

```
wine.isnull().sum().sum()
```

Out[19]:

```
12
```

In [20]:

```
wine.fillna(wine.mean())
```

Out[20]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates
0	7.4	0.700	0.000000	1.9	0.076	11.0	34.0	0.99780	3.51	0.56
1	7.8	0.880	0.000000	2.6	0.098	25.0	67.0	0.99680	3.20	0.68
2	7.8	0.760	0.040000	2.3	0.092	15.0	54.0	0.99700	3.26	0.65
3	11.2	0.280	0.560000	1.9	0.075	17.0	60.0	0.99800	3.16	0.58
4	7.4	0.700	0.000000	1.9	0.076	11.0	34.0	0.99780	3.51	0.56
...
1138	6.3	0.510	0.130000	2.3	0.076	29.0	40.0	0.99574	3.42	0.75
1139	6.8	0.620	0.080000	1.9	0.068	28.0	38.0	0.99651	3.42	0.82
1140	6.2	0.600	0.080000	2.0	0.090	32.0	44.0	0.99490	3.45	0.58
1141	5.9	0.550	0.100000	2.2	0.062	39.0	51.0	0.99512	3.52	0.76
1142	5.9	0.645	0.268561	2.0	0.075	32.0	44.0	0.99547	3.57	0.71

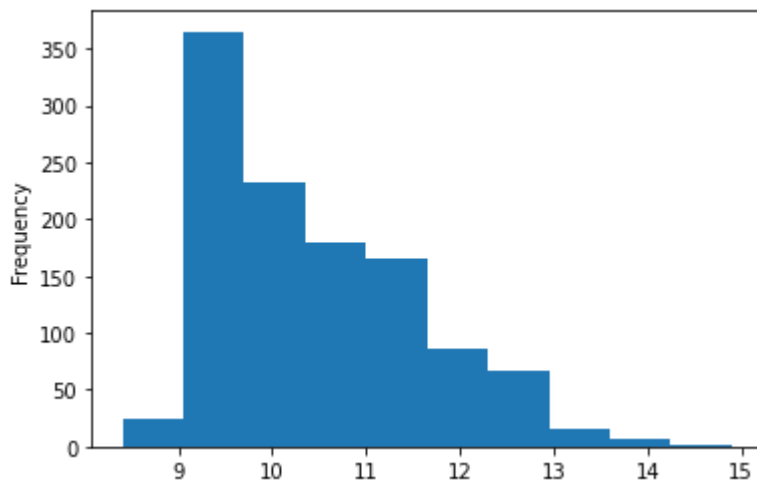
1143 rows × 13 columns

In [27]:

```
al = wine['alcohol']
al.plot(kind='hist')
```

Out[27]:

<AxesSubplot:ylabel='Frequency'>



In [31]:

```
wine.rename(columns={"residual sugar": "sugar"}, inplace=True)
```

In [32]:

```
result_sorted = wine.sort_values(by = ["quality", "alcohol"], inplace = False, ascending=False)
result_sorted.head()
```

Out[32]:

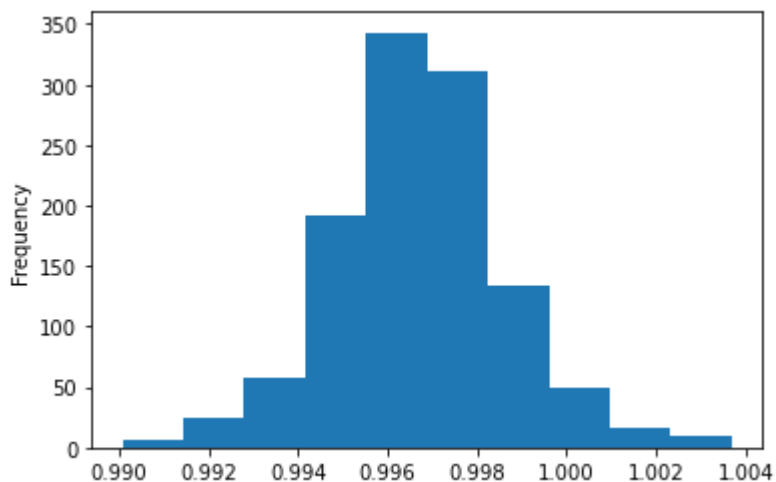
	fixed acidity	volatile acidity	citric acid	sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
419	5.0	0.42	0.24	2.0	0.060	19.0	50.0	0.99170	3.72	0.74	14
321	11.3	0.62	0.67	5.2	0.086	6.0	19.0	0.99880	3.22	0.69	13
793	7.9	0.54	0.34	2.5	0.076	8.0	17.0	0.99235	3.20	0.72	13
271	5.6	0.85	0.05	1.4	0.045	12.0	88.0	0.99240	3.56	0.82	12
190	7.9	0.35	0.46	3.6	0.078	15.0	37.0	0.99730	3.35	0.86	12

In [33]:

```
den = wine['density']
den.plot(kind='hist')
```

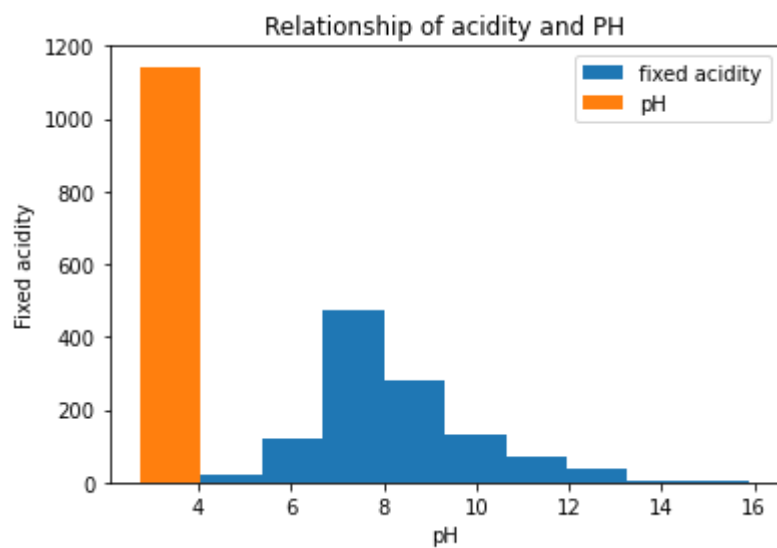
Out[33]:

<AxesSubplot:ylabel='Frequency'>



In [39]:

```
new_wine = wine[['fixed acidity', 'pH']]
new_wine.plot(kind='hist')
plt.xlabel('pH')
plt.ylabel('Fixed acidity')
plt.title('Relationship of acidity and PH')
plt.show()
```



In []: