Concrete Compresive strength analysis and Prediction Model

Data Visualization : Matplotlib

Importing Libraries

In [120]:

```
# Read Data
import numpy as np
import pandas as pd

# Visualization
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
plt.rcParams['figure.figsize']=(10,8)

# style
plt.style.use("fivethirtyeight")
sns.set_style("darkgrid")
```

Read Data

In [121]:

```
# Import first 5 rows
df = pd.read_csv("concrete.csv")
df.head()
```

Out[121]:

	cement	slag	ash	water	superplastic	coarseagg	fineagg	age	strength
0	540.0	0.0	0.0	162.0	2.5	1040.0	676.0	28	79.99
1	540.0	0.0	0.0	162.0	2.5	1055.0	676.0	28	61.89
2	332.5	142.5	0.0	228.0	0.0	932.0	594.0	270	40.27
3	332.5	142.5	0.0	228.0	0.0	932.0	594.0	365	41.05
4	198.6	132.4	0.0	192.0	0.0	978.4	825.5	360	44.30

In [3]:

```
# shape of data ie Rows x Columns
df.shape
```

Out[3]:

```
(1030, 9)
```

In [11]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1030 entries, 0 to 1029
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
```

```
 0   cement         1030 non-null   float64
 1   slag           1030 non-null   float64
 2   ash            1030 non-null   float64
 3   water          1030 non-null   float64
 4   superplastic   1030 non-null   float64
 5   coarseagg      1030 non-null   float64
 6   fineagg        1030 non-null   float64
 7   age            1030 non-null   int64
 8   strength       1030 non-null   float64
dtypes: float64(8), int64(1)
memory usage: 72.5 KB
```

Missing Data

In [4]:

```
# checking for null values
df.isnull().sum()
```

Out[4]:

```
cement          0
slag            0
ash             0
water           0
superplastic    0
coarseagg       0
fineagg         0
age             0
strength        0
dtype: int64
```

In [14]:

```
df.describe()
```

Out[14]:

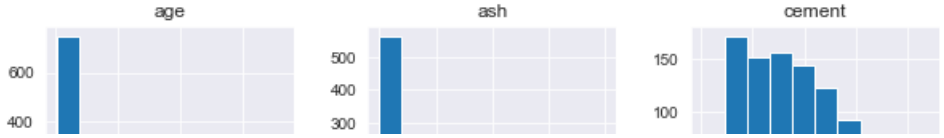
	cement	slag	ash	water	superplastic	coarseagg	fineagg	age	strength
count	1030.000000	1030.000000	1030.000000	1030.000000	1030.000000	1030.000000	1030.000000	1030.000000	1030.000000
mean	281.167864	73.895825	54.188350	181.567282	6.204660	972.918932	773.580485	45.662136	35.817961
std	104.506364	86.279342	63.997004	21.354219	5.973841	77.753954	80.175980	63.169912	16.705742
min	102.000000	0.000000	0.000000	121.800000	0.000000	801.000000	594.000000	1.000000	2.330000
25%	192.375000	0.000000	0.000000	164.900000	0.000000	932.000000	730.950000	7.000000	23.710000
50%	272.900000	22.000000	0.000000	185.000000	6.400000	968.000000	779.500000	28.000000	34.445000
75%	350.000000	142.950000	118.300000	192.000000	10.200000	1029.400000	824.000000	56.000000	46.135000
max	540.000000	359.400000	200.100000	247.000000	32.200000	1145.000000	992.600000	365.000000	82.600000

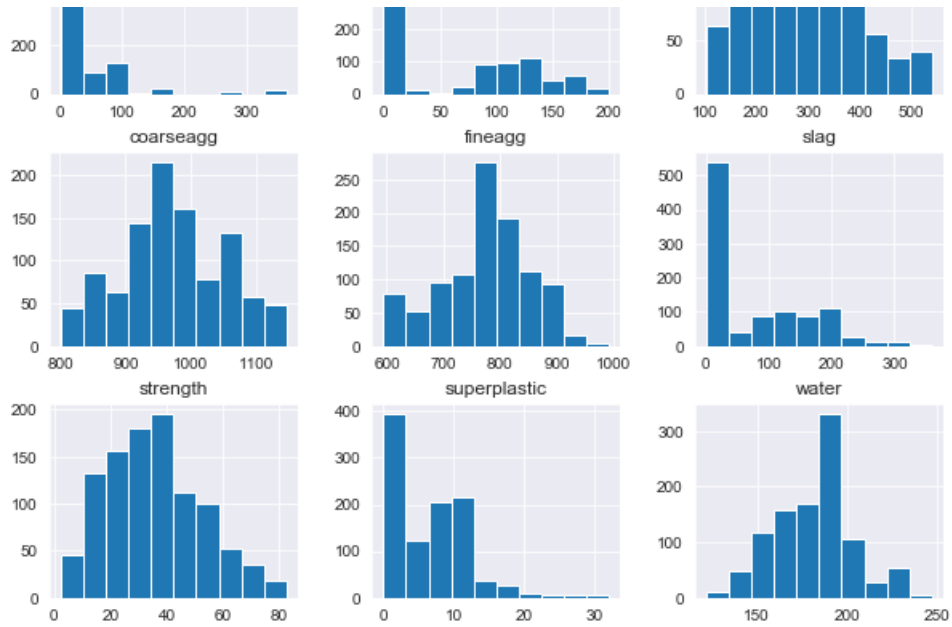
Data Visualization

Univariate Analysis

In [44]:

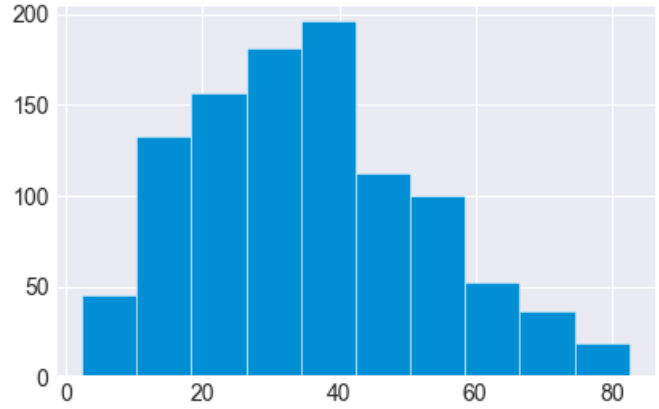
```
# Histrogram
df.hist(figsize = (10,8))
plt.show()
```





In [70]:

```
plt.hist(df.strength)
plt.show()
```



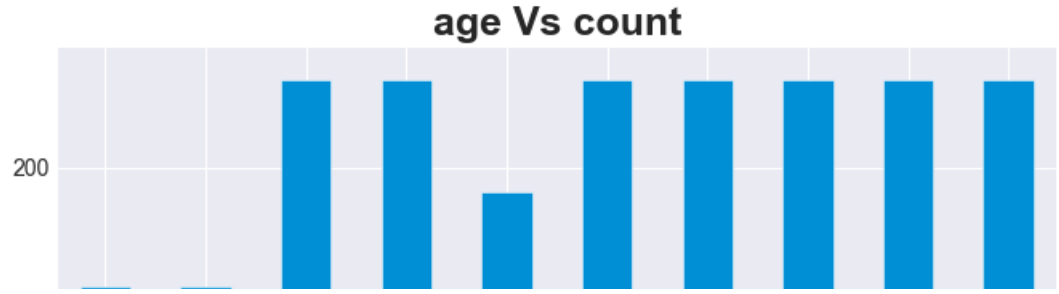
- Target variable 'strength' is right skewed and imbalanced data.

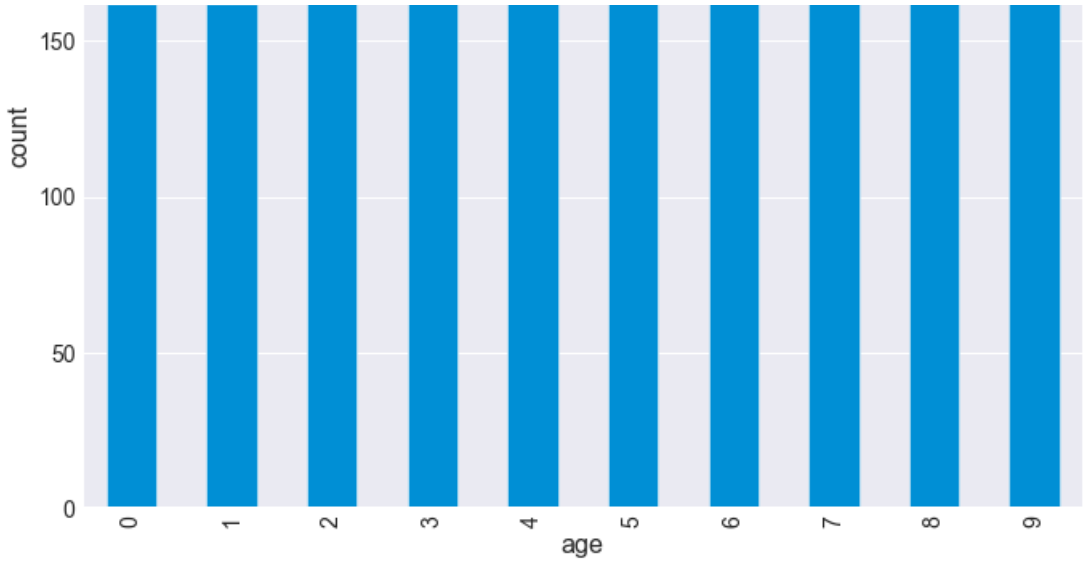
In [88]:

```
data = df['water'].head(10)
data.plot.bar()

plt.xlabel('age', fontsize = 16)
plt.ylabel('count', fontsize = 16)

plt.title('age Vs count', fontsize = 25, fontweight = 'bold')
```

Out[88]:

```
Text(0.5, 1.0, 'age Vs count')
```



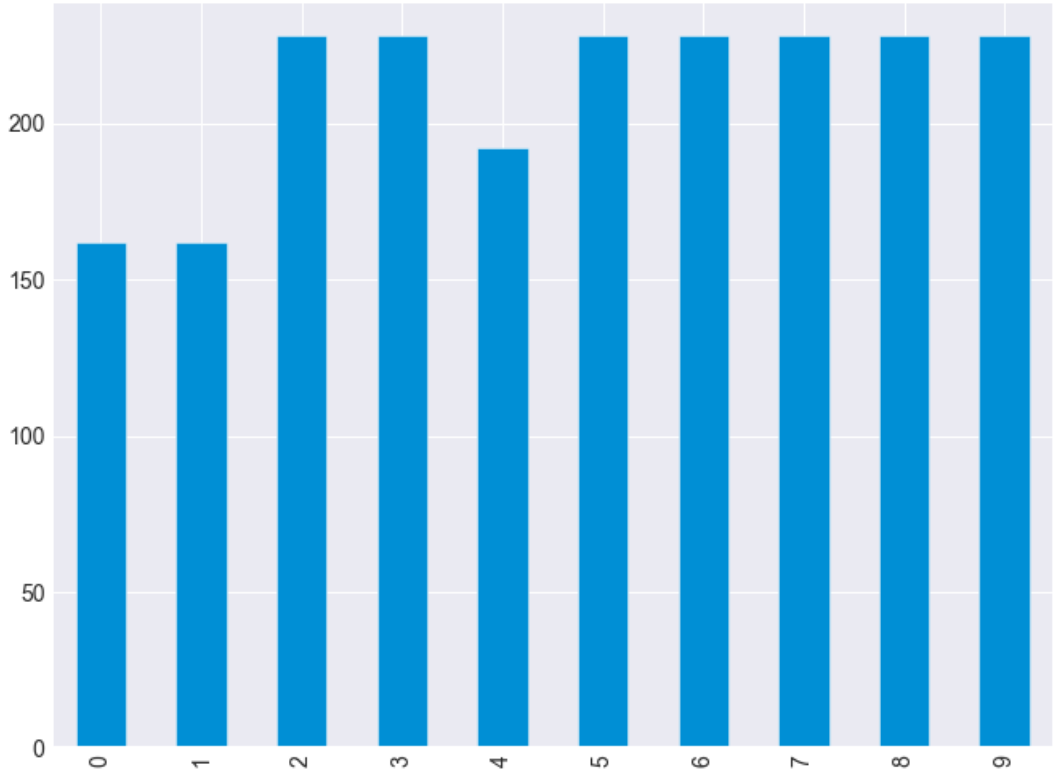


In [89]:

```
x = df['cement'].head(10)
y = df['strength'].head(10)
data.plot.bar(x,y)
```

Out[89]:

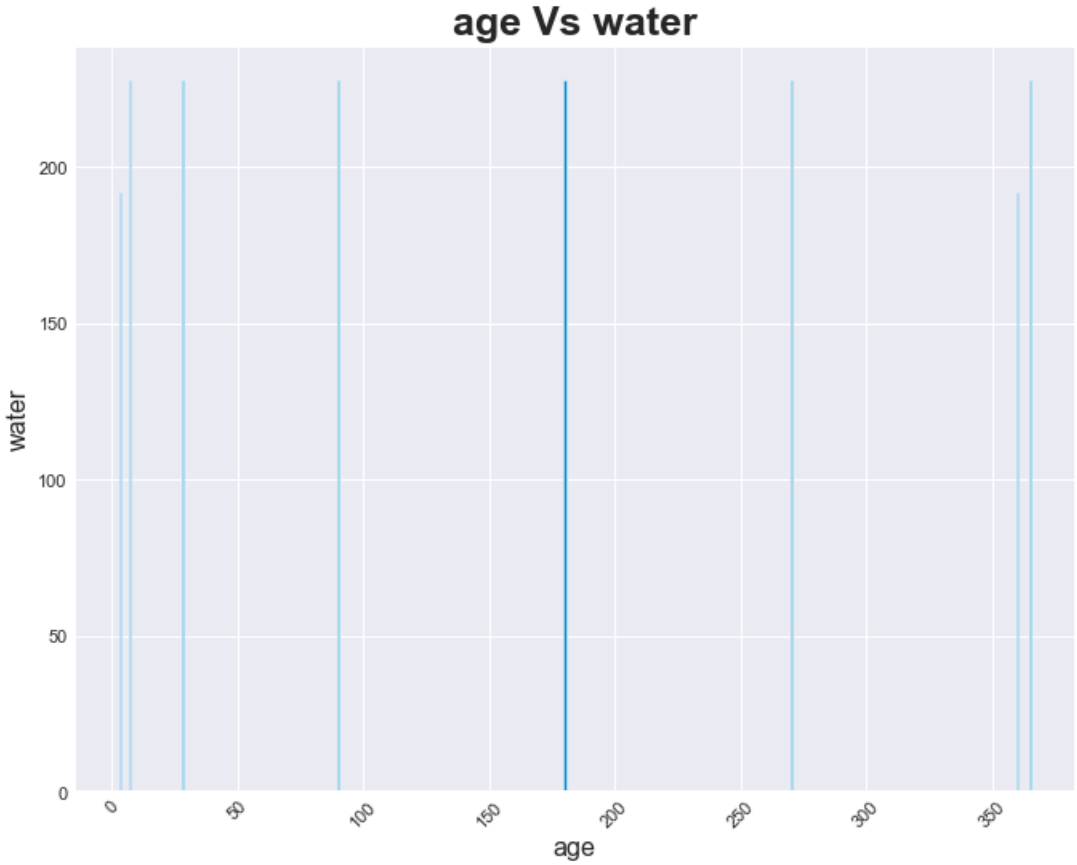
```
<matplotlib.axes._subplots.AxesSubplot at 0x20287f50b38>
```



In [77]:

```
plt.figure(figsize=(10,8))

x = df['age'].head(30)
data = df['water'].head(30)

plt.bar(x,data);

#plt.legend()

plt.xlabel('age', fontsize = 16)
plt.ylabel('water', fontsize = 16)
```

```
plt.title('age Vs water', fontsize = 25, fontweight = 'bold')

plt.xticks(rotation = 45, fontsize = 11)
plt.yticks(fontsize = 11)

plt.show()
```



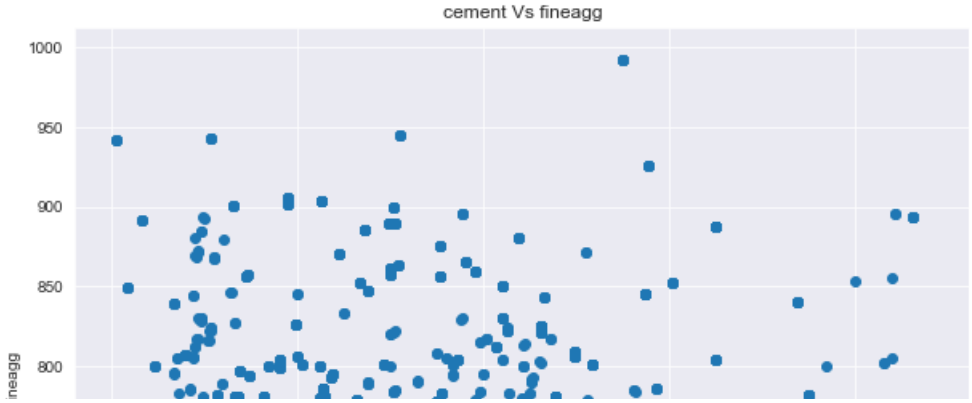
Bivariate Analysis

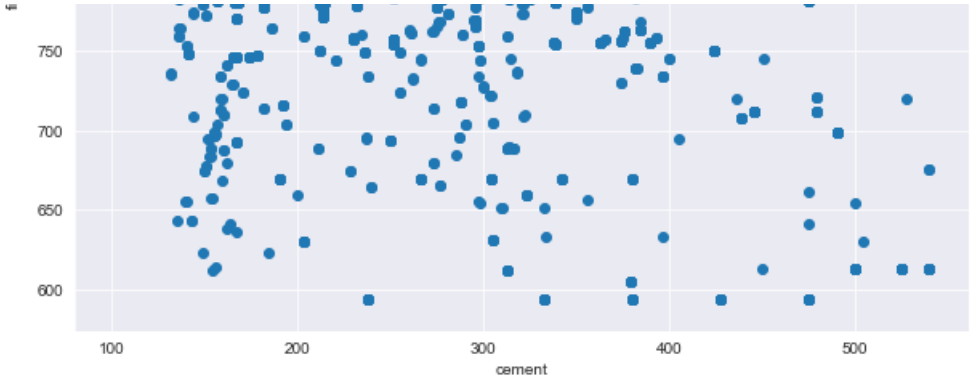
In [21]:

```
# Scatter plot
plt.figure(figsize=(10,8))
plt.scatter(df['cement'], df['fineagg'])

plt.xlabel('cement')
plt.ylabel('fineagg')

plt.title('cement Vs fineagg')
```

Out[21]:

```
Text(0.5, 1.0, 'cement Vs fineagg')
```



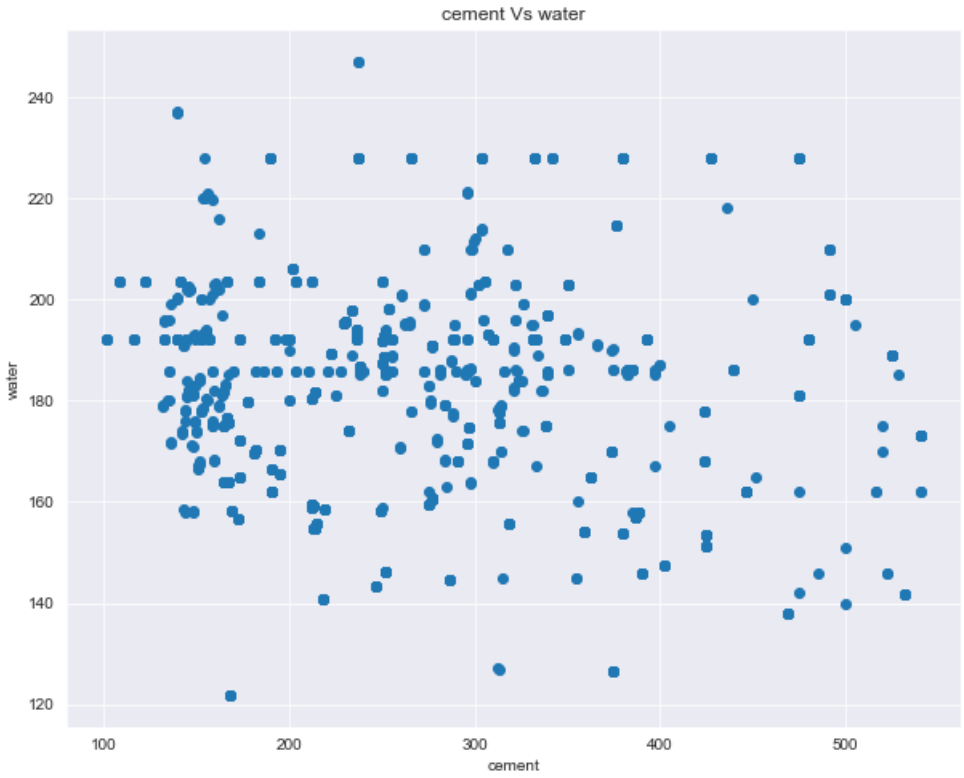


In [25]:

```
plt.figure(figsize=(10,8))
plt.scatter(df['cement'],df['water'])

plt.xlabel('cement')
plt.ylabel('water')

plt.title('cement Vs water')
```

Out[25]:

```
Text(0.5, 1.0, 'cement Vs water')
```

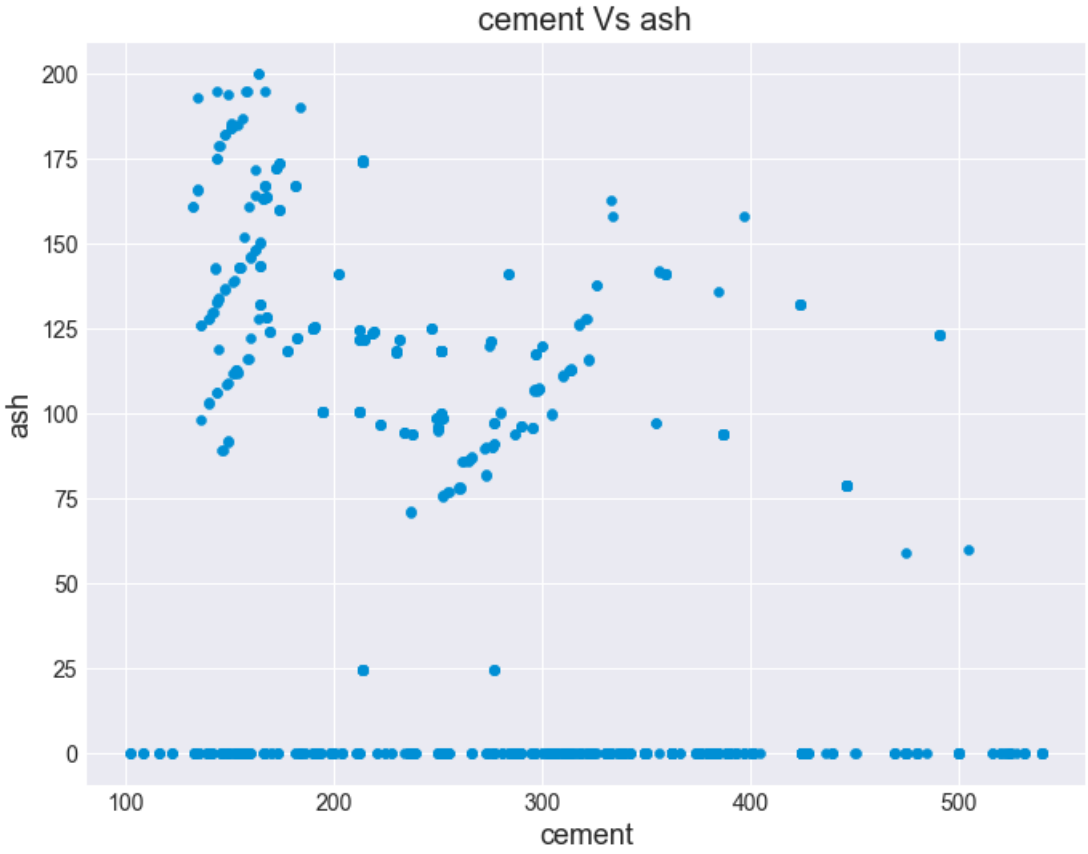


In [131]:

```
plt.figure(figsize=(10,8))
plt.scatter(df['cement'],df['ash'])

plt.xlabel('cement', fontsize=18)
plt.ylabel('ash', fontsize=18)

plt.title('cement Vs ash')
```

Out[131]:

```
Text(0.5, 1.0, 'cement Vs ash')
```

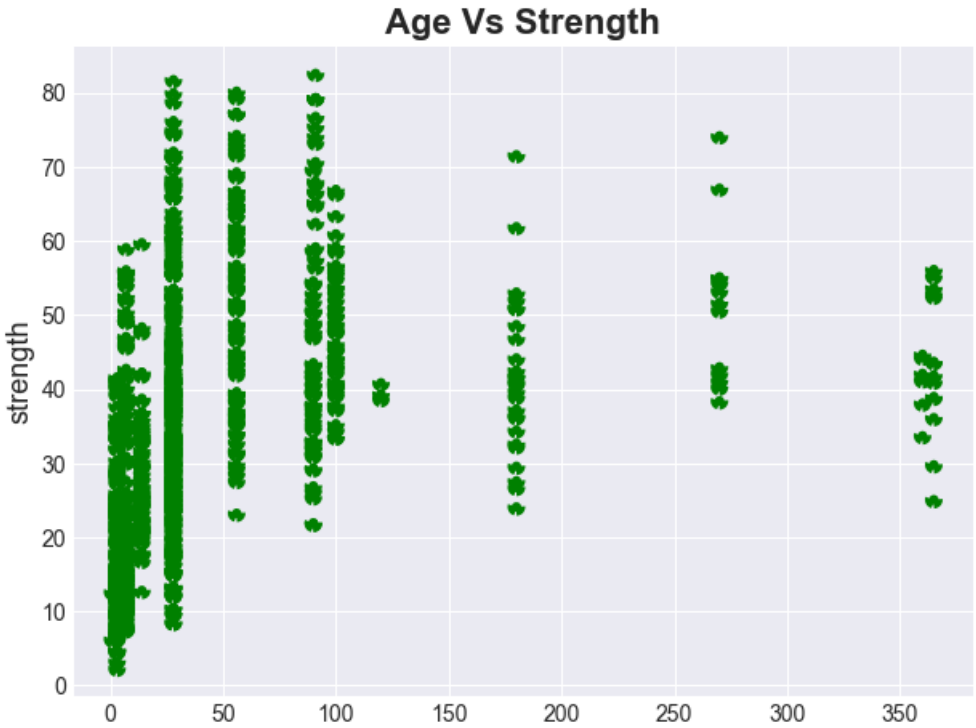


In [133]:

```
plt.figure(figsize=(9,7))
x = df['age']
y = df['strength']
plt.scatter(df['age'], df['strength'], color='green', linewidth=5, linestyle='dotted')

plt.xlabel('age', fontsize=18)
plt.ylabel('strength', fontsize=18)

plt.title('Age Vs Strength', fontsize=22, fontweight='bold')
```

Out[133]:

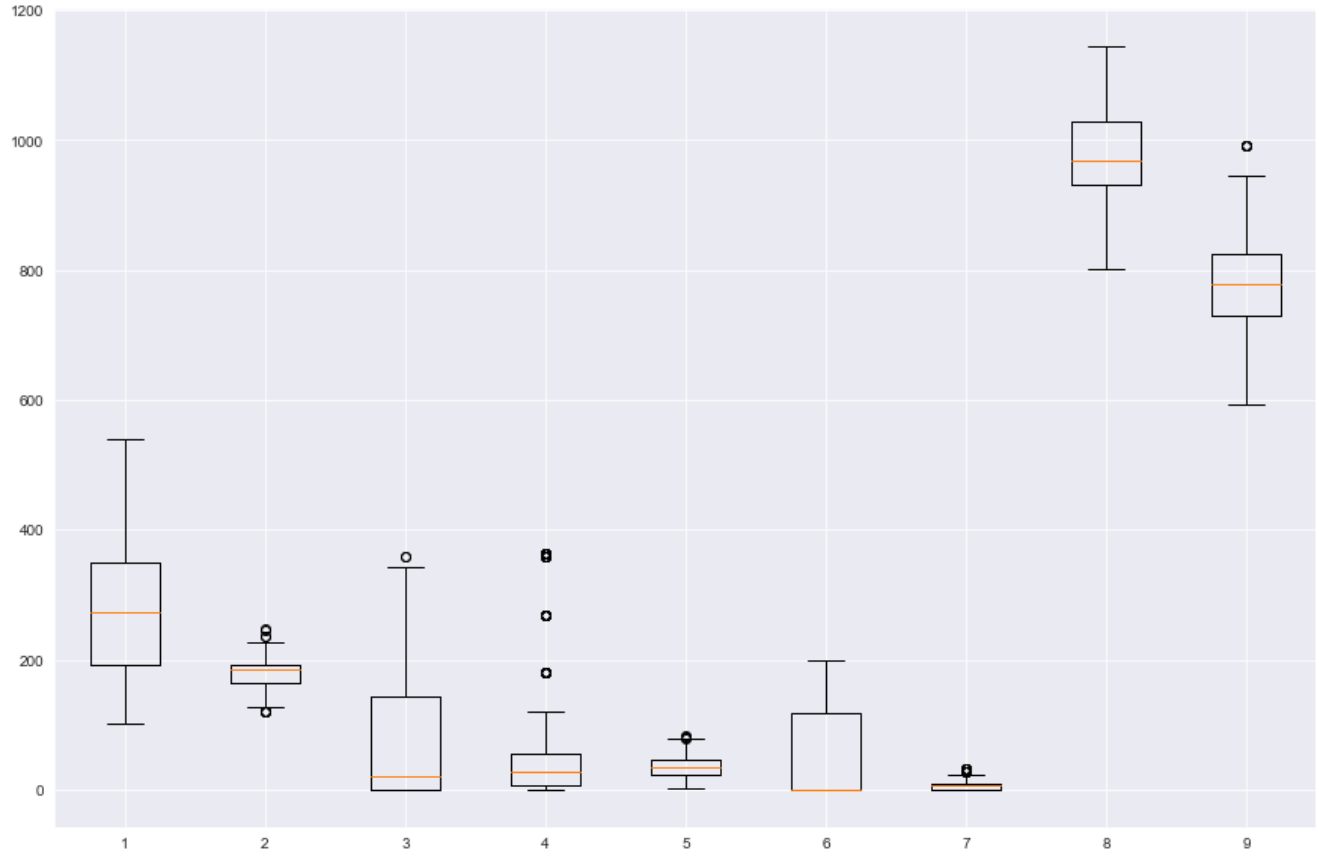
```
Text(0.5, 1.0, 'Age Vs Strength')
```



age

In [43]:

```
# Box Plot
plt.figure(figsize=(15,10))
plt.boxplot([df['cement'],df['water'],df['slag'],df['age'],df['strength'],df['ash'],df['superplast
c'],
                                                df['coarseagg'],df['fineagg']])
plt.show()
```



- Observe that we have outliers in water, slag, age, strength, superplastic, fineagg

---

Data Visualization : Seaborn

Univariate Analysis

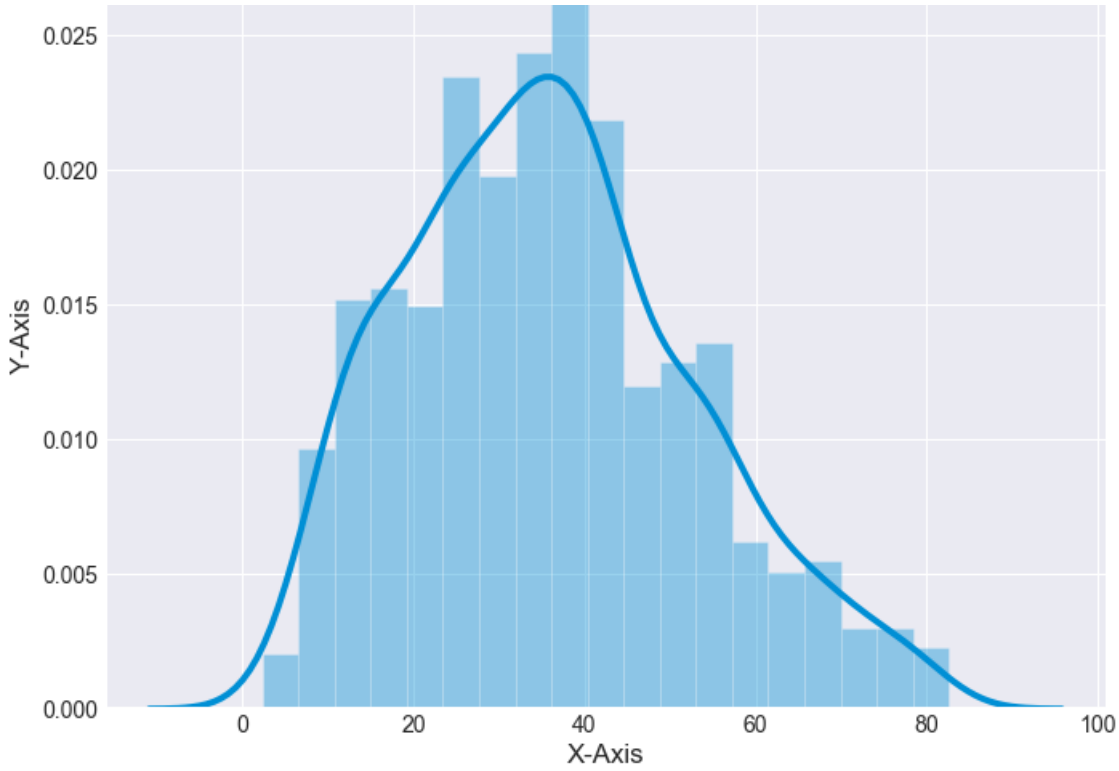
Histogram

In [93]:

```
sns.distplot(df.strength, kde=True)
plt.xlabel('X-Axis')
plt.ylabel('Y-Axis')
```

Out[93]:

```
Text(0, 0.5, 'Y-Axis')
```



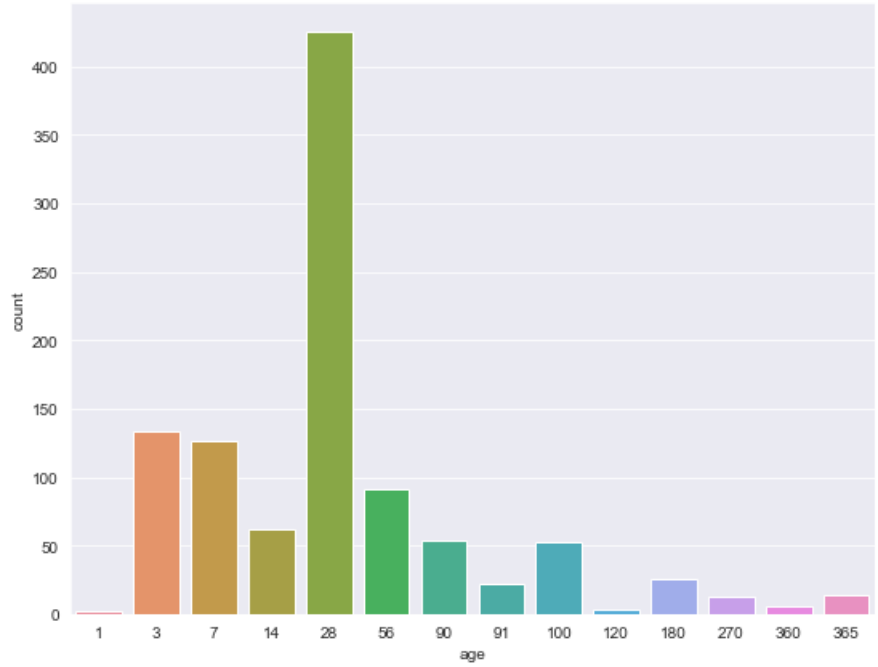
Count Plot

In [19]:

```
plt.figure(figsize=(9,7))
sns.countplot('age', data=df)
```

Out[19]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x202dae86550>
```



Missing Values

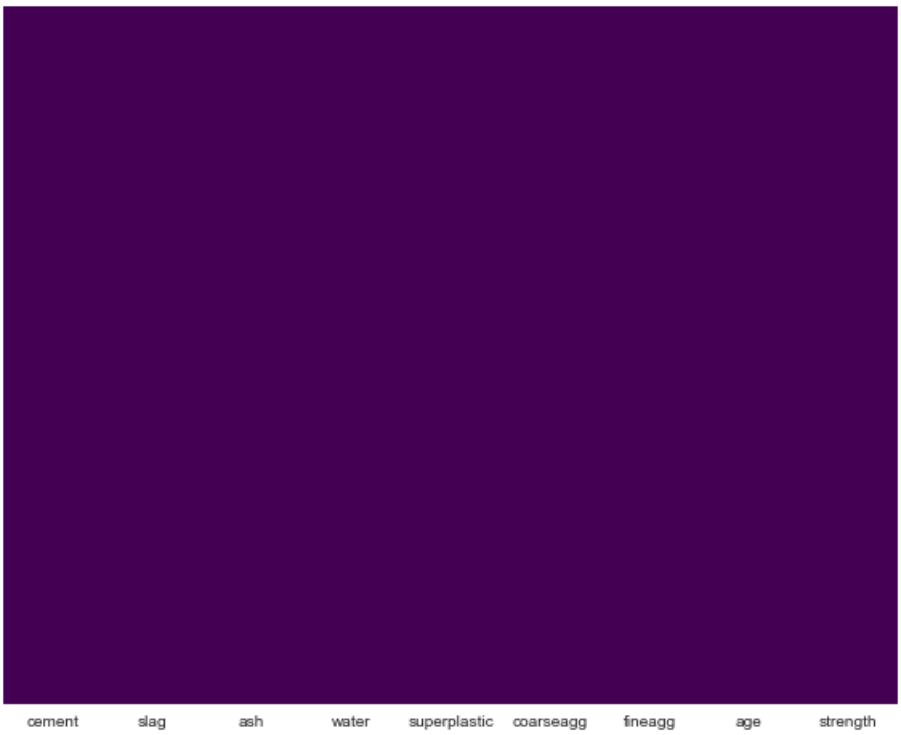
In [13]:

```
# Missing value representation by Heatmap
plt.figure(figsize=(10,8))
```

```
plt.figure(figsize=(10,8))
sns.heatmap(df.isnull(), yticklabels=False, cbar=False, cmap='viridis')
```

Out[13]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x202da541198>
```

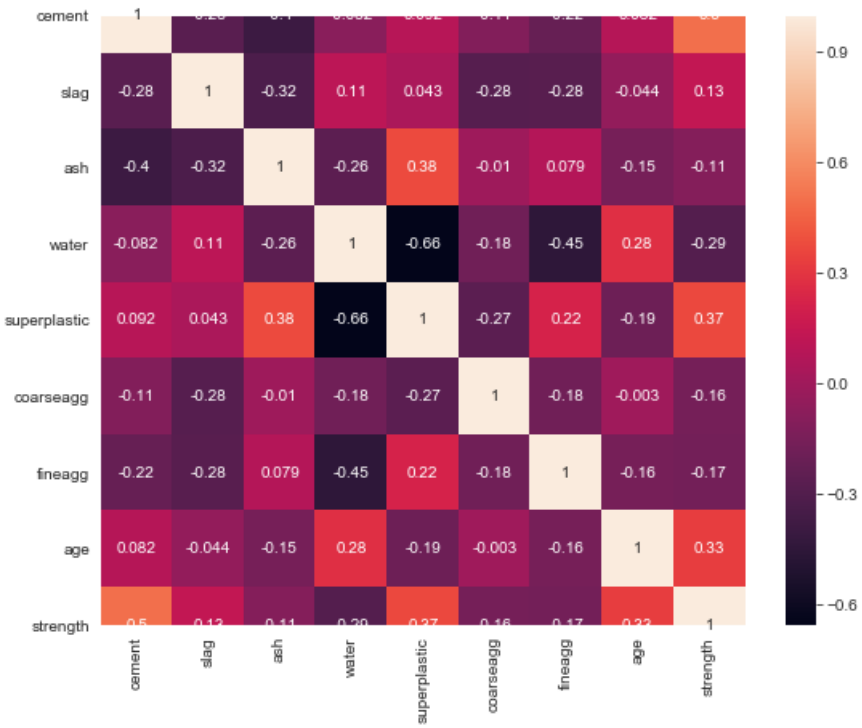


In [16]:

```
plt.figure(figsize=(9,7))
sns.heatmap(df.corr(), annot=True)
```

Out[16]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x202da7849e8>
```



Pair Plot

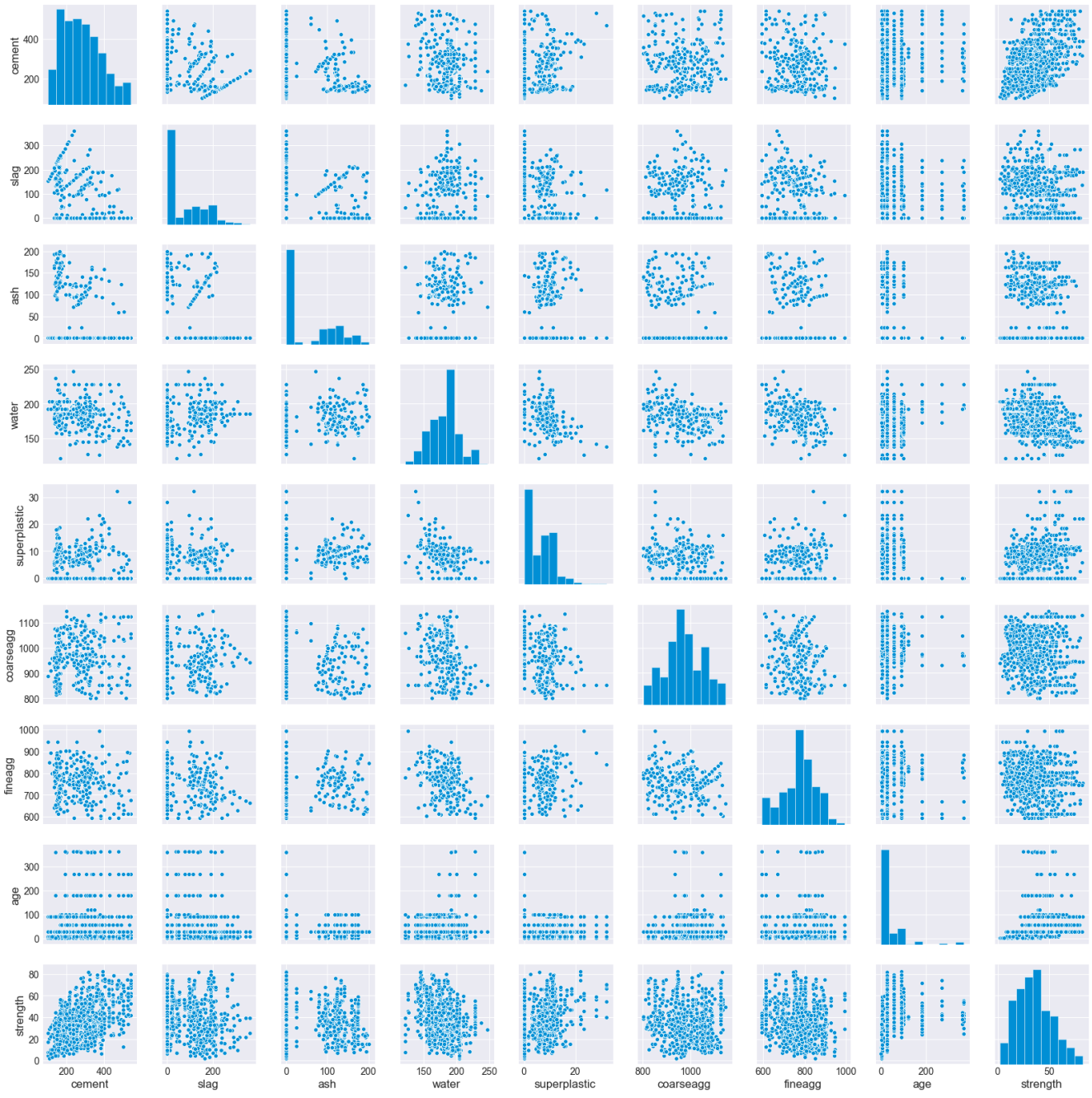
- to represent relation between each feature

In [112]:

```
sns.pairplot(df)
```

Out[112]:

```
<seaborn.axisgrid.PairGrid at 0x20298da9dd8>
```



Box Plot

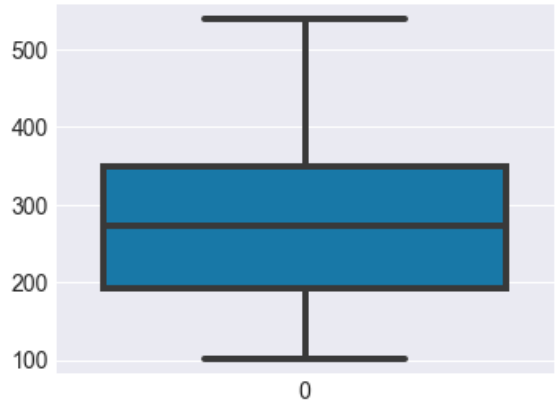
- to find out outliers

In [99]:

```
plt.figure(figsize=(5,4))
sns.boxplot(data=df.cement, palette='winter')
```

Out[99]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x202909c5c50>
```

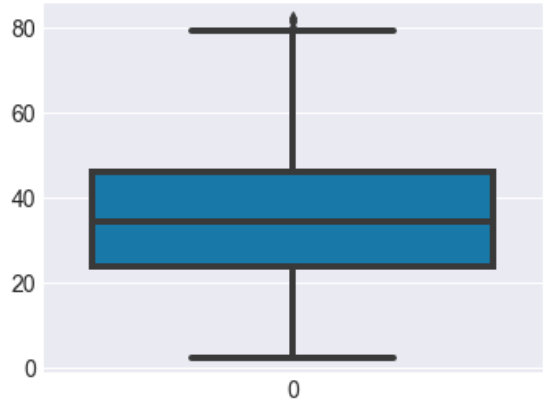


In [100]:

```
plt.figure(figsize=(5,4))
sns.boxplot(data=df.strength, palette='winter')
```

Out[100]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x20290572be0>
```



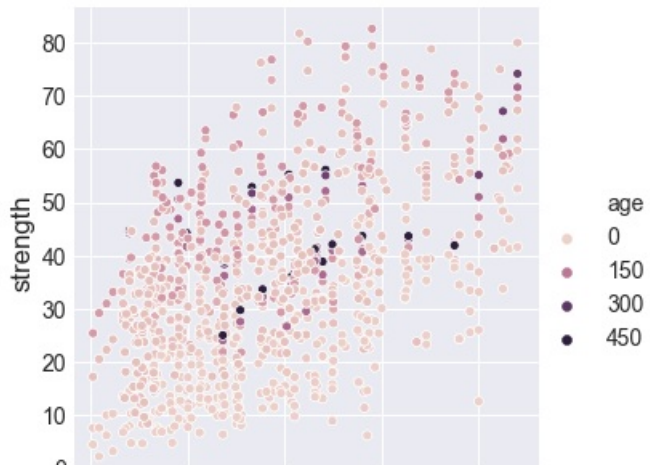
Relational Plot

In [110]:

```
sns.relplot(x='cement', y='strength', data=df, hue="age")
```

Out[110]:

```
<seaborn.axisgrid.FacetGrid at 0x2029344b1d0>
```



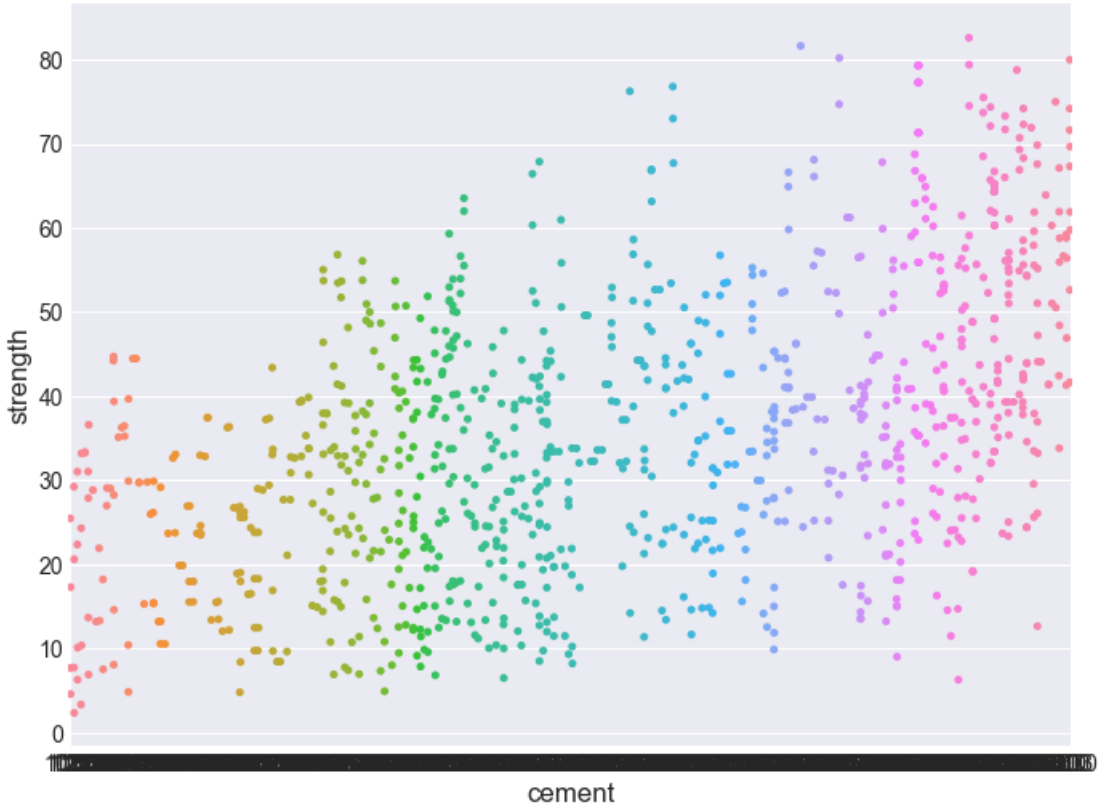
Strip Plot

In [113]:

```
sns.stripplot(x='cement', y='strength', data=df)
```

Out[113]:

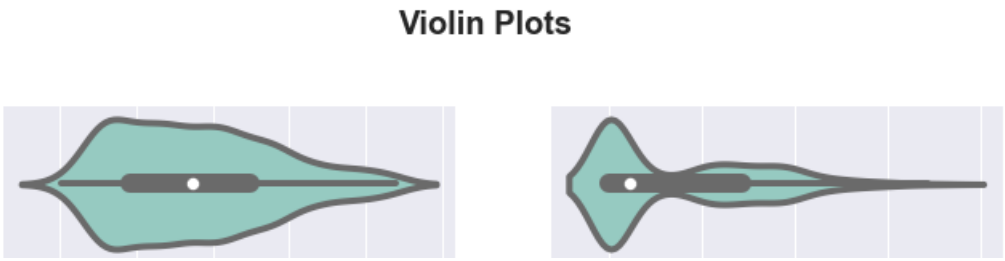
```
<matplotlib.axes._subplots.AxesSubplot at 0x2029cb21c50>
```

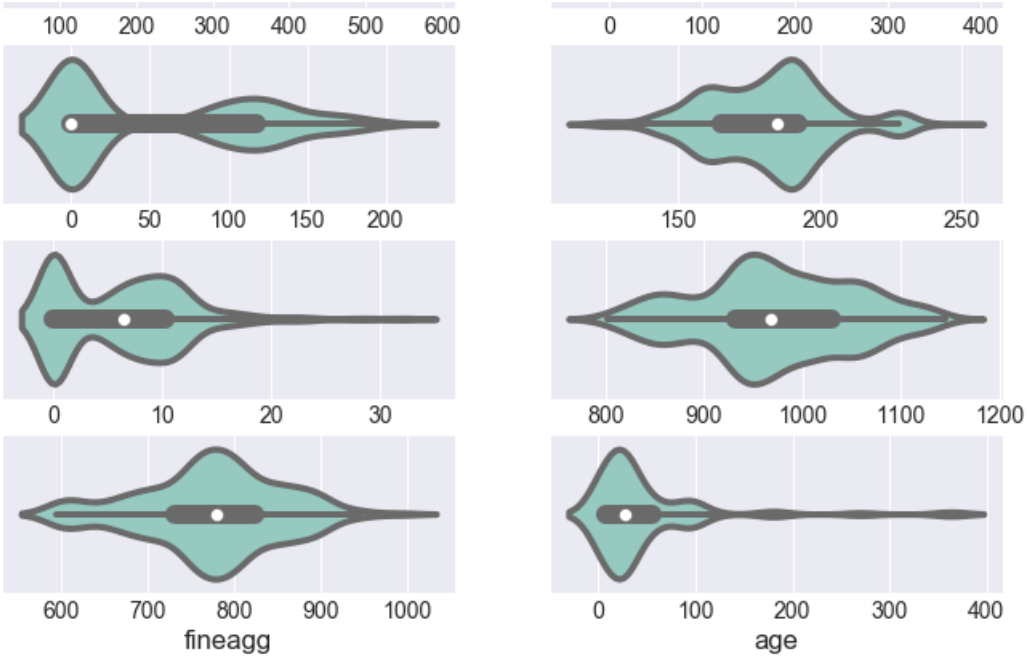


Voilin Plot

In [129]:

```
fig,ax = plt.subplots(nrows=4, ncols=2, figsize=(10,8))
plt.suptitle('Violin Plots', fontsize=20, fontweight = 'bold')
sns.violinplot(x="cement", data=df, ax=ax[0,0], palette='Set3')
sns.violinplot(x="slag", data=df, ax=ax[0,1], palette='Set3')
sns.violinplot (x='ash', data=df, ax=ax[1,0], palette='Set3')
sns.violinplot(x='water', data=df, ax=ax[1,1], palette='Set3')
sns.violinplot(x="superplastic", data=df, ax=ax[2,0], palette='Set3')
sns.violinplot(x="coarseagg", data=df, ax=ax[2,1], palette='Set3')
sns.violinplot (x ='fineagg', data=df, ax=ax[3,0], palette='Set3')
sns.violinplot(x='age', data=df, ax=ax[3,1], palette='Set3')
plt.show()
```





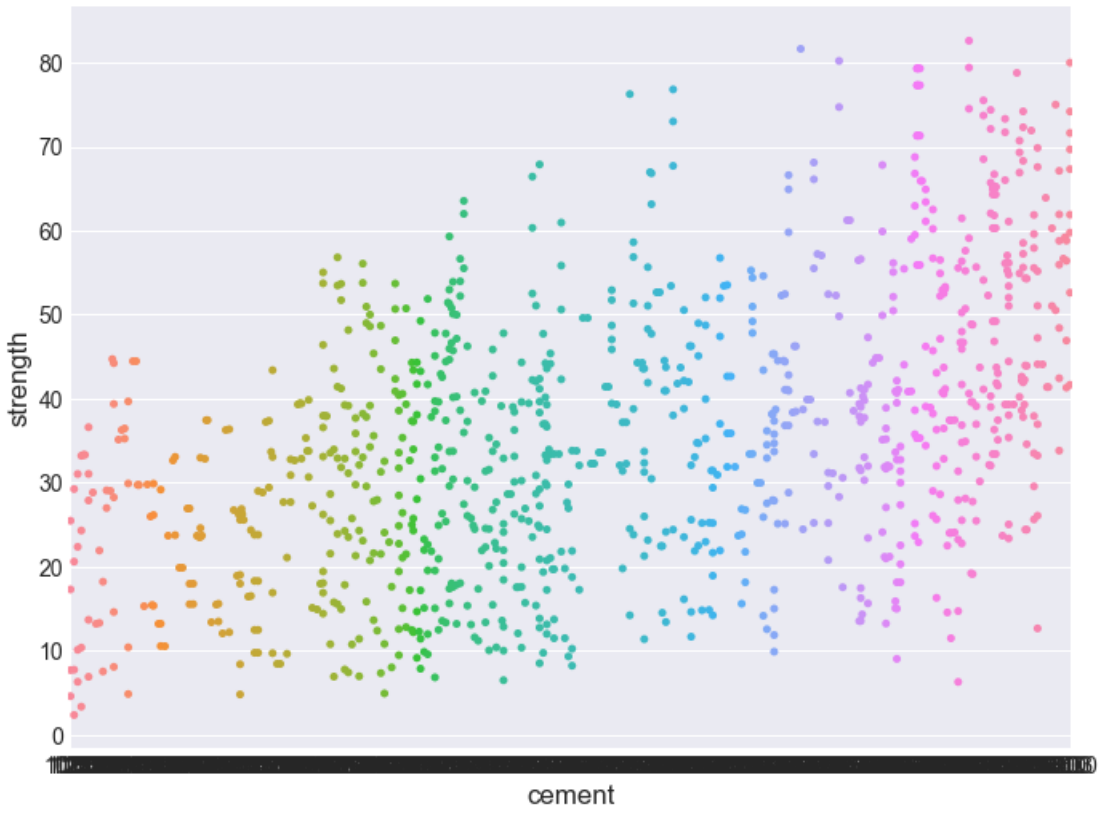
Swarm Plot

In [118]:

```
sns.swarmplot(x='cement', y='strength', data=df)
```

Out[118]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2029f7ed7b8>
```



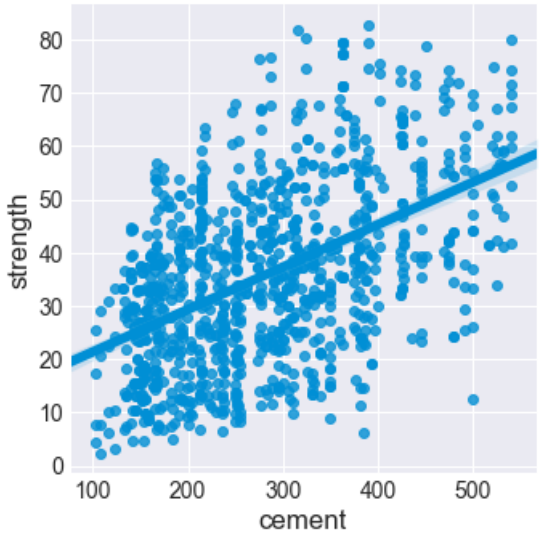
Regression Plot

In [119]:

```
sns.lmplot(x='cement', y='strength', data=df)
```

Out[119]:

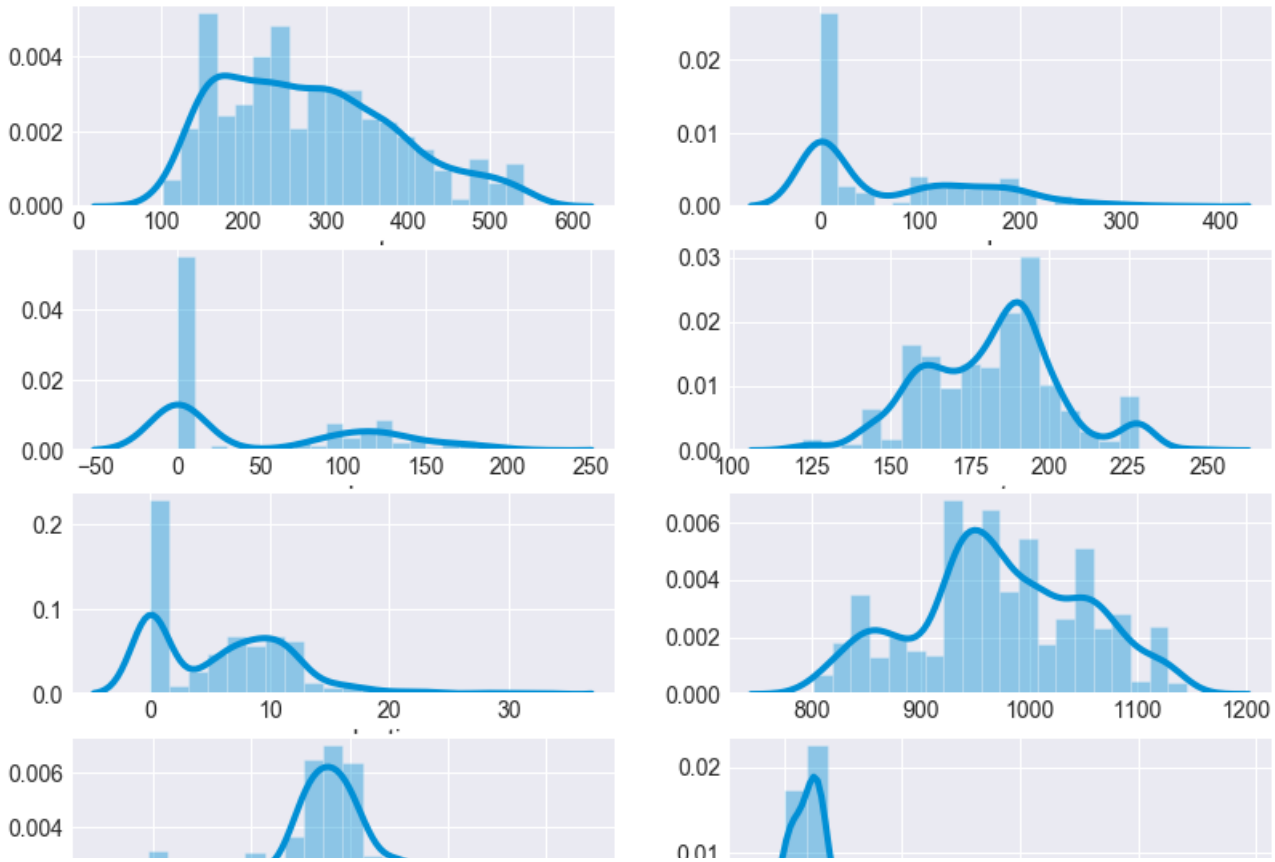
```
<seaborn.axisgrid.FacetGrid at 0x202a0384160>
```

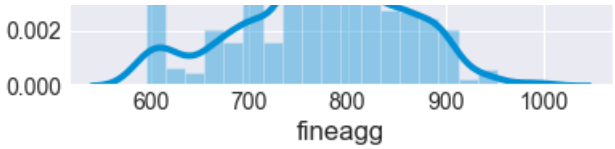


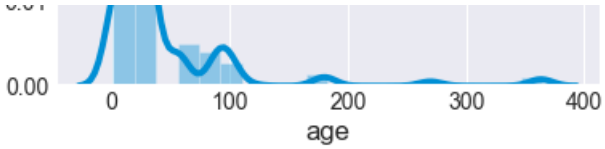
Dist Plot

In [124]:

```
fig, ax = plt.subplots(4,2, figsize=(12,10))
sns.distplot(df.cement, bins = 20, ax=ax[0,0])
sns.distplot(df.slag, bins = 20, ax=ax[0,1])
sns.distplot(df.ash, bins = 20, ax=ax[1,0])
sns.distplot(df.water, bins = 20, ax=ax[1,1])
sns.distplot(df.superplastic, bins = 20, ax=ax[2,0])
sns.distplot(df.coarseagg, bins = 20, ax=ax[2,1])
sns.distplot(df.fineagg, bins = 20, ax=ax[3,0])
sns.distplot(df.age, bins = 20, ax=ax[3,1])
plt.show()
```




0.002
0.000
600
700
800
900
1000
fineagg


0.00
0
100
200
300
400
age