In [1]:

```
import requests
import pandas as pd
url="https://raw.githubusercontent.com/realpython/materials/master/pandas-gradebook-project
res=requests.get(url,allow_redirects=True)
with open('roster.csv','wb') as file:
    file.write(res.content)
roster=pd.read_csv('roster.csv',converters={"NetID": str.lower, "Email Address": str.lower}
```

In [2]:

```
roster.head()
```

Out[2]:

	ID	Name	NetID	Email Address	Section
0	1234567	Barrera Jr., Woody	wxb12345	woody.barrera_jr@univ.edu	1
1	2345678	Lambert, Malaika	mxl12345	malaika.lambert@univ.edu	2
2	3456789	Joyce, Traci	txj12345	traci.joyce@univ.edu	1
3	4567890	Flower, John Gregg	jgf12345	john.g.2.flower@univ.edu	3
4	8406082	Johnson, Stacy Michael	smj00936	stacy.johnson@univ.edu	2

In [3]:

```
roster.tail()
```

Out[3]:

	ID	Name	NetID	Email Address	Section
145	5227994	Johnson, Paul Mark	pmj37756	paul.johnson@univ.edu	3
146	9804556	Lee, Danielle Stephanie	dsl24347	danielle.lee@univ.edu	3
147	1783442	Edwards, Nicole	nxe44872	nicole.edwards@univ.edu	3
148	6345758	Reyes, Bailey	bxr62103	bailey.reyes@univ.edu	2
149	5566719	Walls, Joyce	jxw53347	joyce.walls@univ.edu	1

In [4]:

```
url="https://raw.githubusercontent.com/realpython/materials/master/pandas-gradebook-project
res=requests.get(url,allow_redirects=True)
with open('hm_exams_grades.csv','wb') as file:
    file.write(res.content)
hm_exams_grades=pd.read_csv('hm_exams_grades.csv')
```

In [5]:

```
hm_exams_grades.head()
```

Out[5]:

	First Name	Last Name	SID	Homework 1	Homework 1 - Max Points	Homework 1 - Submission Time	Homework 2	Homework 2 - Max Points
0	Aaron	Lester	axl60952	68.0	80	2019-08-29 08:56:02-07:00	74	80
1	Adam	Cooper	amc28428	80.0	80	2019-08-29 08:56:02-07:00	78	80
2	Alec	Curry	axc64717	69.0	80	2019-08-29 08:56:02-07:00	76	80
3	Alexander	Rodriguez	akr14831	50.0	80	2019-08-29 08:56:02-07:00	54	80
4	Amber	Daniels	axd11293	54.0	80	2019-08-29 08:56:02-07:00	57	80

5 rows × 42 columns

In [6]:

```
url1="https://raw.githubusercontent.com/realpython/materials/master/pandas-gradebook-projec
url2="https://raw.githubusercontent.com/realpython/materials/master/pandas-gradebook-projec
url3="https://raw.githubusercontent.com/realpython/materials/master/pandas-gradebook-projec
url4="https://raw.githubusercontent.com/realpython/materials/master/pandas-gradebook-projec
url5="https://raw.githubusercontent.com/realpython/materials/master/pandas-gradebook-projec
res1=requests.get(url1,allow_redirects=True)
with open('quiz_1.csv','wb') as file:
    file.write(res1.content)
quiz_1=pd.read_csv('quiz_1.csv')

res2=requests.get(url2,allow_redirects=True)
with open('quiz_2.csv','wb') as file:
    file.write(res2.content)
quiz_2=pd.read_csv('quiz_2.csv')

res3=requests.get(url3,allow_redirects=True)
with open('quiz_3.csv','wb') as file:
    file.write(res3.content)
quiz_3=pd.read_csv('quiz_3.csv')

res4=requests.get(url4,allow_redirects=True)
with open('quiz_4.csv','wb') as file:
    file.write(res4.content)
quiz_4=pd.read_csv('quiz_4.csv')

res5=requests.get(url5,allow_redirects=True)
with open('quiz_5.csv','wb') as file:
    file.write(res5.content)
quiz_5=pd.read_csv('quiz_5.csv')
```

In [7]:

```
quiz_1.rename(columns={"Grade":"Quiz 1"},inplace=True)
quiz_1.drop(labels=['Last Name','First Name'],axis=1,inplace=True)
quiz_1
```

Out[7]:

	Email	Quiz 1
0	richard.bennett@univ.edu	10
1	timothy.parker@univ.edu	9
2	carol.reyes@univ.edu	5
3	brooke.powers@univ.edu	6
4	michael.taylor@univ.edu	5
...	...	...
145	jeffrey.perez@univ.edu	4
146	angela.dunlap@univ.edu	6
147	richard.elliott@univ.edu	6
148	donna.nguyen@univ.edu	7
149	timothy.ramirez@univ.edu	4

150 rows × 2 columns

In [8]:

```
quiz_2.rename(columns={"Grade":"Quiz 2"},inplace=True)
quiz_2.drop(labels=['Last Name','First Name'],axis=1,inplace=True)
# quiz_2.set_index('Email',inplace=True)
quiz_2
```

Out[8]:

	Email	Quiz 2
0	daisy.anderson@univ.edu	6
1	john.g.2.flower@univ.edu	8
2	samantha.smith@univ.edu	7
3	amy.white@univ.edu	14
4	desiree.delgado@univ.edu	7
...	...	...
145	carol.reyes@univ.edu	15
146	cameron.dennis@univ.edu	6
147	joseph.young@univ.edu	14
148	richard.elliott@univ.edu	13
149	felicia.simpson@univ.edu	7

150 rows × 2 columns

In [9]:

```
quiz_3.rename(columns={"Grade":"Quiz 3"},inplace=True)
quiz_3.drop(labels=['Last Name','First Name'],axis=1,inplace=True)
# quiz_3.set_index('Email',inplace=True)
quiz_3.head()
```

Out[9]:

	Email	Quiz 3
0	brandon.thomas@univ.edu	13
1	john.g.2.flower@univ.edu	8
2	taylor.hernandez@univ.edu	14
3	tamara.warren@univ.edu	16
4	felicia.simpson@univ.edu	13

In [10]:

```
quiz_4.rename(columns={"Grade":"Quiz 4"},inplace=True)
quiz_4.drop(labels=['Last Name','First Name'],axis=1,inplace=True)
# quiz_4.set_index('Email',inplace=True)
quiz_4.head()
```

Out[10]:

	Email	Quiz 4
0	nicole.patterson@univ.edu	13
1	benjamin.graham@univ.edu	6
2	russell.flores@univ.edu	9
3	kimberly.moore@univ.edu	8
4	amy.adams@univ.edu	8

In [11]:

```
quiz_5.rename(columns={"Grade":"Quiz 5"},inplace=True)
quiz_5.drop(labels=['Last Name','First Name'],axis=1,inplace=True)
# quiz_5.set_index('Email',inplace=True)
quiz_5.head()
```

Out[11]:

	Email	Quiz 5
0	rachel.dennis@univ.edu	11
1	courtney.wolf@univ.edu	9
2	aaron.lester@univ.edu	8
3	ashley.martin@univ.edu	11
4	jessica.garza@univ.edu	10

In [12]:

```
roster
```

Out[12]:

	ID	Name	NetID	Email Address	Section
0	1234567	Barrera Jr., Woody	wxb12345	woody.barrera_jr@univ.edu	1
1	2345678	Lambert, Malaika	mxl12345	malaika.lambert@univ.edu	2
2	3456789	Joyce, Traci	txj12345	traci.joyce@univ.edu	1
3	4567890	Flower, John Gregg	jgf12345	john.g.2.flower@univ.edu	3
4	8406082	Johnson, Stacy Michael	smj00936	stacy.johnson@univ.edu	2
...	...	...	...	...	...
145	5227994	Johnson, Paul Mark	pmj37756	paul.johnson@univ.edu	3
146	9804556	Lee, Danielle Stephanie	dsl24347	danielle.lee@univ.edu	3
147	1783442	Edwards, Nicole	nxe44872	nicole.edwards@univ.edu	3
148	6345758	Reyes, Bailey	bxr62103	bailey.reyes@univ.edu	2
149	5566719	Walls, Joyce	jxw53347	joyce.walls@univ.edu	1

150 rows × 5 columns

In [13]:

```
roster.drop(axis=1,labels=['ID','Name'],inplace=True)
```

In [14]:

```
roster
```

Out[14]:

	NetID	Email Address	Section
0	wxb12345	woody.barrera_jr@univ.edu	1
1	mxl12345	malaika.lambert@univ.edu	2
2	txj12345	traci.joyce@univ.edu	1
3	jgf12345	john.g.2.flower@univ.edu	3
4	smj00936	stacy.johnson@univ.edu	2
...	...	...	...
145	pmj37756	paul.johnson@univ.edu	3
146	dsl24347	danielle.lee@univ.edu	3
147	nxe44872	nicole.edwards@univ.edu	3
148	bxr62103	bailey.reyes@univ.edu	2
149	jxw53347	joyce.walls@univ.edu	1

150 rows × 3 columns

In [15]:

```
roster.set_index('NetID',inplace=True)
```

In [16]:

```
roster.head()
```

Out[16]:

NetID	Email Address	Section
wxb12345	woody.barrera_jr@univ.edu	1
mxl12345	malaika.lambert@univ.edu	2
txj12345	traci.joyce@univ.edu	1
jgf12345	john.g.2.flower@univ.edu	3
smj00936	stacy.johnson@univ.edu	2

In [17]:

```
hm_exams_grades.head()
```

Out[17]:

	First Name	Last Name	SID	Homework 1	Homework 1 - Max Points	Homework 1 - Submission Time	Homework 2	Homework 2 - Max Points
0	Aaron	Lester	axl60952	68.0	80	2019-08-29 08:56:02-07:00	74	80
1	Adam	Cooper	amc28428	80.0	80	2019-08-29 08:56:02-07:00	78	80
2	Alec	Curry	axc64717	69.0	80	2019-08-29 08:56:02-07:00	76	80
3	Alexander	Rodriguez	akr14831	50.0	80	2019-08-29 08:56:02-07:00	54	80
4	Amber	Daniels	axd11293	54.0	80	2019-08-29 08:56:02-07:00	57	80

5 rows × 42 columns

In [18]:

```
hm_exams_grades.drop(axis=1,labels=['First Name','Last Name'],inplace=True)
```

In [19]:

```
hm_exams_grades.set_index('SID',inplace=True)
hm_exams_grades.head()
```

Out[19]:

SID	Homework 1	Homework 1 - Max Points	Homework 1 - Submission Time	Homework 2	Homework 2 - Max Points	Homework 2 - Submission Time	Homework 3
axl60952	68.0	80	2019-08-29 08:56:02-07:00	74	80	2019-09-05 08:56:02-07:00	77
amc28428	80.0	80	2019-08-29 08:56:02-07:00	78	80	2019-09-05 08:56:02-07:00	78
axc64717	69.0	80	2019-08-29 08:56:02-07:00	76	80	2019-09-05 08:56:02-07:00	66
akr14831	50.0	80	2019-08-29 08:56:02-07:00	54	80	2019-09-05 08:56:02-07:00	74
axd11293	54.0	80	2019-08-29 08:56:02-07:00	57	80	2019-09-05 08:56:02-07:00	77

5 rows × 39 columns

In [20]:

```
quiz_grades1=pd.merge(quiz_1,quiz_2,on='Email')
quiz_grades2=pd.merge(quiz_grades1,quiz_3,on='Email')
quiz_grades3=pd.merge(quiz_grades2,quiz_4,on='Email')
quiz_grades_final=pd.merge(quiz_grades3,quiz_5,on='Email')
quiz_grades_final
```

Out[20]:

	Email	Quiz 1	Quiz 2	Quiz 3	Quiz 4	Quiz 5
0	richard.bennett@univ.edu	10	6	9	8	10
1	timothy.parker@univ.edu	9	14	13	14	10
2	carol.reyes@univ.edu	5	15	8	14	6
3	brooke.powers@univ.edu	6	10	17	10	8
4	michael.taylor@univ.edu	5	15	13	12	5
...	...	...	...	...	...	...
145	jeffrey.perez@univ.edu	4	7	12	12	9
146	angela.dunlap@univ.edu	6	11	11	11	6
147	richard.elliott@univ.edu	6	13	17	11	12
148	donna.nguyen@univ.edu	7	12	14	9	4
149	timothy.ramirez@univ.edu	4	8	13	6	4

150 rows × 6 columns

In [21]:

```
quiz_grades_final.set_index('Email',inplace=True)
quiz_grades_final.head()
```

Out[21]:

Email	Quiz 1	Quiz 2	Quiz 3	Quiz 4	Quiz 5
richard.bennett@univ.edu	10	6	9	8	10
timothy.parker@univ.edu	9	14	13	14	10
carol.reyes@univ.edu	5	15	8	14	6
brooke.powers@univ.edu	6	10	17	10	8
michael.taylor@univ.edu	5	15	13	12	5

In [22]:

```
roster.head(1)
```

Out[22]:

NetID	Email Address	Section
wxb12345	woody.barrera_jr@univ.edu	1

In [23]:

```
hm_exams_grades.head(1)
```

Out[23]:

SID	Homework 1	Homework 1 - Max Points	Homework 1 - Submission Time	Homework 2	Homework 2 - Max Points	Homework 2 - Submission Time	Homework 3
axl60952	68.0	80	2019-08-29 08:56:02-07:00	74	80	2019-09-05 08:56:02-07:00	77

1 rows × 39 columns

In [24]:

```
final_data=pd.merge(roster,hm_exams_grades,left_index=True,right_index=True)
final_data.head()
```

Out[24]:

	Email Address	Section	Homework 1	Homework 1 - Max Points	Homework 1 - Submission Time	Homework 2	H
wxb12345	woody.barrera_jr@univ.edu	1	55.0	80	2019-08-29 08:56:02-07:00	62	
mxl12345	malaika.lambert@univ.edu	2	63.0	80	2019-08-29 08:56:02-07:00	57	
txj12345	traci.joyce@univ.edu	1	NaN	80	2019-08-29 08:56:02-07:00	77	
jgf12345	john.g.2.flower@univ.edu	3	69.0	80	2019-08-29 08:56:02-07:00	52	
smj00936	stacy.johnson@univ.edu	2	74.0	80	2019-08-29 08:56:02-07:00	55	

5 rows × 41 columns

In [25]:

```
final_data=pd.merge(final_data,quiz_grades_final,left_on="Email Address", right_index=True)
final_data.head()
```

Out[25]:

	Email Address	Section	Homework 1	Homework 1 - Max Points	Homework 1 - Submission Time	Homework 2	H
wxb12345	woody.barrera_jr@univ.edu	1	55.0	80	2019-08-29 08:56:02-07:00	62	
mxl12345	malaika.lambert@univ.edu	2	63.0	80	2019-08-29 08:56:02-07:00	57	
txj12345	traci.joyce@univ.edu	1	NaN	80	2019-08-29 08:56:02-07:00	77	
jgf12345	john.g.2.flower@univ.edu	3	69.0	80	2019-08-29 08:56:02-07:00	52	
smj00936	stacy.johnson@univ.edu	2	74.0	80	2019-08-29 08:56:02-07:00	55	

5 rows × 46 columns

```
# now data is ready to analyse
```

fill in nan values

In [26]:

```
final_data=final_data.fillna(0)
final_data.head()
```

Out[26]:

	Email Address	Section	Homework 1	Homework 1 - Max Points	Homework 1 - Submission Time	Homework 2	H
wxb12345	woody.barrera_jr@univ.edu	1	55.0	80	2019-08-29 08:56:02-07:00	62	
mxl12345	malaika.lambert@univ.edu	2	63.0	80	2019-08-29 08:56:02-07:00	57	
txj12345	traci.joyce@univ.edu	1	0.0	80	2019-08-29 08:56:02-07:00	77	
jgf12345	john.g.2.flower@univ.edu	3	69.0	80	2019-08-29 08:56:02-07:00	52	
smj00936	stacy.johnson@univ.edu	2	74.0	80	2019-08-29 08:56:02-07:00	55	

5 rows × 46 columns

calculating the total score

In [27]:

```
n_exams = 3
for n in range(1, n_exams + 1):
    final_data[f"Exam {n} Score"] = (
        final_data[f"Exam {n}"] / final_data[f"Exam {n} - Max Points"]
    )
```

Calculating the homework score

In [28]:

```
homework_scores = final_data.filter(regex=r"^Homework \d\d?$", axis=1)
homework_max_points = final_data.filter(regex=r"^Homework \d\d? - Max Points$", axis=1)
```

In [29]:

```
sum_of_hw_scores = homework_scores.sum(axis=1)
sum_of_hw_max = homework_max_points.sum(axis=1)
final_data["Total Homework"] = sum_of_hw_scores /sum_of_hw_max
```

In [30]:

```
hw_scores=final_data.filter(regex=r"^Hw \d$",axis=1)
```

In [31]:

```
homework_max_points.head()
```

Out[31]:

	Homework 1 - Max Points	Homework 2 - Max Points	Homework 3 - Max Points	Homework 4 - Max Points	Homework 5 - Max Points	Homework 6 - Max Points	Homework 7 - Max Points	H
wxb12345	80	80	80	100	70	90	50	
mxl12345	80	80	80	100	70	90	50	
txj12345	80	80	80	100	70	90	50	
jgf12345	80	80	80	100	70	90	50	
smj00936	80	80	80	100	70	90	50	

In [32]:

```
homework_scores.head()
```

Out[32]:

	Homework 1	Homework 2	Homework 3	Homework 4	Homework 5	Homework 6	Homework 7	H
wxb12345	55.0	62	73	83	68	87	36	
mxl12345	63.0	57	78	83	61	88	42	
txj12345	0.0	77	58	91	66	78	30	
jgf12345	69.0	52	64	75	61	83	41	
smj00936	74.0	55	60	67	46	85	43	

In [33]:

```
homework_scores.columns
```

Out[33]:

```
Index(['Homework 1', 'Homework 2', 'Homework 3', 'Homework 4', 'Homework 5',
       'Homework 6', 'Homework 7', 'Homework 8', 'Homework 9', 'Homework 1
0'],
      dtype='object')
```

In [34]:

```
hw_max_renamed = homework_max_points.set_axis(homework_scores.columns, axis=1)
```

In [35]:

```
hw_max_renamed.head()
```

Out[35]:

	Homework 1	Homework 2	Homework 3	Homework 4	Homework 5	Homework 6	Homework 7	H
wxb12345	80	80	80	100	70	90	50	
mxl12345	80	80	80	100	70	90	50	
txj12345	80	80	80	100	70	90	50	
jgf12345	80	80	80	100	70	90	50	
smj00936	80	80	80	100	70	90	50	

In [36]:

```
average_hw_scores = (homework_scores / hw_max_renamed).sum(axis=1)
average_hw_scores.head()
```

Out[36]:

```
wxb12345    7.994048
mxl12345    8.189444
txj12345    7.859405
jgf12345    7.657103
smj00936    7.877421
dtype: float64
```

In [37]:

```
homework_scores.shape[1]
```

Out[37]:

```
10
```

In [38]:

```
final_data["Average Homework"] = average_hw_scores / homework_scores.shape[1]
```

In [39]:

```
final_data["Homework Score"] = final_data[
    ["Total Homework", "Average Homework"]
].max(axis=1)
```

In [40]:

```
final_data.head()
```

Out[40]:

	Email Address	Section	Homework 1	Homework 1 - Max Points	Homework 1 - Submission Time	Homework 2	H
wxb12345	woody.barrera_jr@univ.edu	1	55.0	80	2019-08-29 08:56:02-07:00	62	
mxl12345	malaika.lambert@univ.edu	2	63.0	80	2019-08-29 08:56:02-07:00	57	
txj12345	traci.joyce@univ.edu	1	0.0	80	2019-08-29 08:56:02-07:00	77	
jgf12345	john.g.2.flower@univ.edu	3	69.0	80	2019-08-29 08:56:02-07:00	52	
smj00936	stacy.johnson@univ.edu	2	74.0	80	2019-08-29 08:56:02-07:00	55	

5 rows × 52 columns

Calculating the quiz csore

In [41]:

```
quiz_scores=final_data.filter(regex=r"^Quiz \d$",axis=1)
quiz_max_points = pd.Series(
    {"Quiz 1": 11, "Quiz 2": 15, "Quiz 3": 17, "Quiz 4": 14, "Quiz 5": 12}
)
sum_quiz_score=quiz_scores.sum(axis=1)
sum_quiz_max_points=quiz_max_points.sum()
final_data["Total quizzes"]=sum_quiz_score/sum_quiz_max_points
avg_quiz_scores = (quiz_scores/quiz_max_points).sum(axis=1)
final_data["Avg quizzes"]=avg_quiz_scores/quiz_scores.shape[1]
final_data["Quiz Score"]=final_data[["Total quizzes","Avg quizzes"]].max(axis=1)
final_data.head()
```

Out[41]:

	Email Address	Section	Homework 1	Homework 1 - Max Points	Homework 1 - Submission Time	Homework 2	H
wxb12345	woody.barrera_jr@univ.edu	1	55.0	80	2019-08-29 08:56:02-07:00	62	
mxl12345	malaika.lambert@univ.edu	2	63.0	80	2019-08-29 08:56:02-07:00	57	
txj12345	traci.joyce@univ.edu	1	0.0	80	2019-08-29 08:56:02-07:00	77	
jgf12345	john.g.2.flower@univ.edu	3	69.0	80	2019-08-29 08:56:02-07:00	52	
smj00936	stacy.johnson@univ.edu	2	74.0	80	2019-08-29 08:56:02-07:00	55	

5 rows × 55 columns

In [42]:

```
weightings = pd.Series(
    {
        "Exam 1 Score": 0.05,
        "Exam 2 Score": 0.1,
        "Exam 3 Score": 0.15,
        "Quiz Score": 0.30,
        "Homework Score": 0.4,
    }
)
```

In [43]:

```
import numpy as np
final_data["Final Score"] = (final_data[weightings.index] * weightings).sum(
    axis=1
)
final_data["Ceiling Score"] = np.ceil(final_data["Final Score"] * 100)
```

In [44]:

```
grades = {
    90: "A",
    80: "B",
    70: "C",
    60: "D",
    0: "F",
}
def grade_mapping(value):
    for key,letter in grades.items():
        if value>=key:
            return letter
```

In [45]:

```
letter_grades = final_data["Ceiling Score"].map(grade_mapping)
final_data["Final Grade"] = pd.Categorical(
    letter_grades, categories=grades.values(), ordered=True
)
```

In [46]:

```
final_data.head()
```

Out[46]:

	Email Address	Section	Homework 1	Homework 1 - Max Points	Homework 1 - Submission Time	Homework 2	H
wxb12345	woody.barrera_jr@univ.edu	1	55.0	80	2019-08-29 08:56:02-07:00	62	
mxl12345	malaika.lambert@univ.edu	2	63.0	80	2019-08-29 08:56:02-07:00	57	
txj12345	traci.joyce@univ.edu	1	0.0	80	2019-08-29 08:56:02-07:00	77	
jgf12345	john.g.2.flower@univ.edu	3	69.0	80	2019-08-29 08:56:02-07:00	52	
smj00936	stacy.johnson@univ.edu	2	74.0	80	2019-08-29 08:56:02-07:00	55	

5 rows × 58 columns

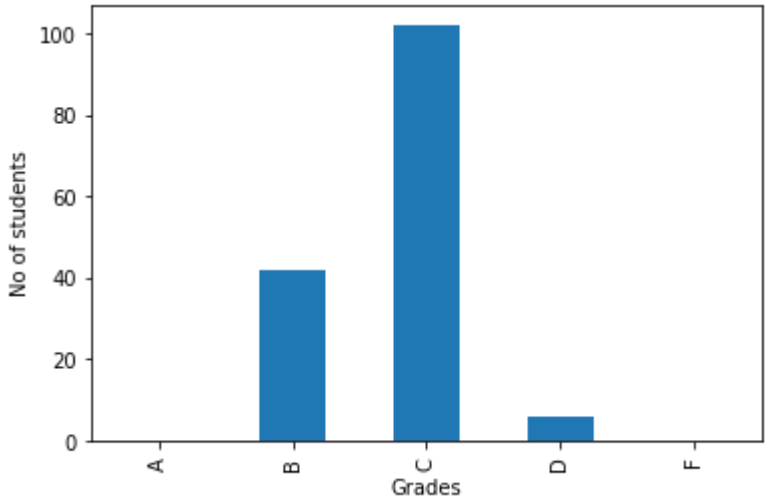
plotting summary statistics

In [47]:

```
import matplotlib.pyplot as plt
import scipy.stats
```

In [48]:

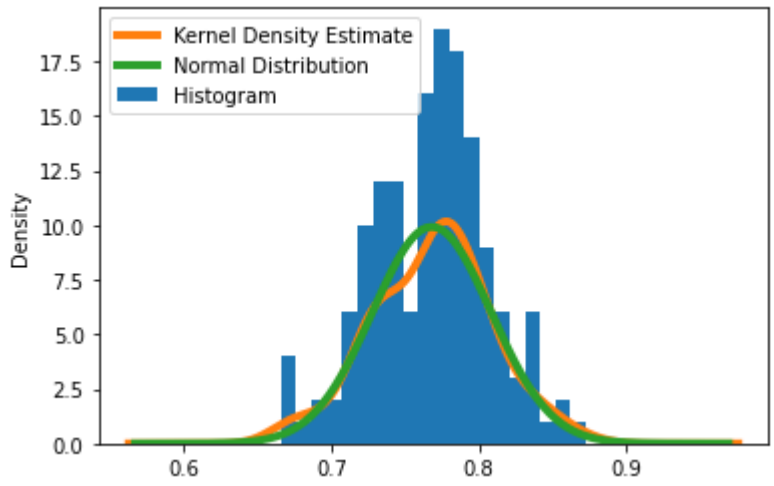
```
grade_counts = final_data["Final Grade"].value_counts().sort_index()
grade_counts.plot.bar()
plt.xlabel("Grades")
plt.ylabel("No of students")
plt.show()
```



In [49]:

```
final_data["Final Score"].plot.hist(bins=20, label="Histogram")
final_data["Final Score"].plot.density(
    linewidth=4, label="Kernel Density Estimate"
)

final_mean = final_data["Final Score"].mean()
final_std = final_data["Final Score"].std()
x = np.linspace(final_mean - 5 * final_std, final_mean + 5 * final_std, 200)
normal_dist = scipy.stats.norm.pdf(x, loc=final_mean, scale=final_std)
plt.plot(x, normal_dist, label="Normal Distribution", linewidth=4)
plt.legend()
plt.show()
```



In [ ]:

```

```