DSBDL Assignment 03 - Descriptive Statistics: Measures of Central Tendency and Variability

Part 1

Provide summary statistics (mean, median, minimum, maximum, standard deviation) for a dataset (age, income etc.) with numeric variables grouped by one of the qualitative (categorical) variable. For example, if your categorical variable is age groups and quantitative variable is income, then provide summary statistics of income grouped by the age groups. Create a list that contains a numeric value for each response to the categorical variable.

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.m
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
ds = pd.read_csv('/content/drive/My Drive/DSBDL/Assignment3/nba.csv')
ds
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0.0	PG	25.0	6-2	180.0	Texas	7730337.0
1	Jae Crowder	Boston Celtics	99.0	SF	25.0	6-6	235.0	Marquette	6796117.0
2	John Holland	Boston Celtics	30.0	SG	27.0	6-5	205.0	Boston University	NaN
3	R.J. Hunter	Boston Celtics	28.0	SG	22.0	6-5	185.0	Georgia State	1148640.0
4	Jonas Jerebko	Boston Celtics	8.0	PF	29.0	6-10	231.0	NaN	5000000.0
...	...	...	...	...	...	...	...	...	...
453	Shelvin Mack	Utah Jazz	8.0	PG	26.0	6-3	203.0	Butler	2433333.0
454	Raul Neto	Utah Jazz	25.0	PG	24.0	6-1	179.0	NaN	900000.0

Next steps: Generate code with ds | View recommended plots

```
ds.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 458 entries, 0 to 457
Data columns (total 9 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   Name      457 non-null    object 
 1   Team      457 non-null    object 
 2   Number    457 non-null    float64
 3   Position  457 non-null    object 
 4   Age       457 non-null    float64
 5   Height    457 non-null    object 
 6   Weight    457 non-null    float64
 7   College   373 non-null    object 
 8   Salary    446 non-null    float64
dtypes: float64(4), object(5)
memory usage: 32.3+ KB
```

```
ds.isna().sum()
```

```
Name         1
Team         1
Number       1
Position     1
Age          1
Height       1
Weight       1
College     85
```

```
Salary      12
dtype: int64
```

```
ds['College'].fillna(ds['College'].mode()[0], inplace = True)
ds['Salary'].fillna(ds['Salary'].mean(), inplace = True)
ds
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0.0	PG	25.0	6-2	180.0	Texas	7.730337e+06
1	Jae Crowder	Boston Celtics	99.0	SF	25.0	6-6	235.0	Marquette	6.796117e+06
2	John Holland	Boston Celtics	30.0	SG	27.0	6-5	205.0	Boston University	4.842684e+06
3	R.J. Hunter	Boston Celtics	28.0	SG	22.0	6-5	185.0	Georgia State	1.148640e+06
4	Jonas Jerebko	Boston Celtics	8.0	PF	29.0	6-10	231.0	Kentucky	5.000000e+06
...	...	...	...	...	...	...	...	...	...
453	Shelvin Mack	Utah Jazz	8.0	PG	26.0	6-3	203.0	Butler	2.433333e+06
454	Raul Neto	Utah Jazz	25.0	PG	24.0	6-1	179.0	Kentucky	9.000000e+05

Next steps: Generate code with ds View recommended plots

```
ds.dropna(inplace = True)
ds
```

	Name	Team	Number	Position	Age	Height	Weight	College	Salary
0	Avery Bradley	Boston Celtics	0.0	PG	25.0	6-2	180.0	Texas	7.730337e+06
1	Jae Crowder	Boston Celtics	99.0	SF	25.0	6-6	235.0	Marquette	6.796117e+06
2	John Holland	Boston Celtics	30.0	SG	27.0	6-5	205.0	Boston University	4.842684e+06
3	R.J. Hunter	Boston Celtics	28.0	SG	22.0	6-5	185.0	Georgia State	1.148640e+06
4	Jonas Jerebko	Boston Celtics	8.0	PF	29.0	6-10	231.0	Kentucky	5.000000e+06
...	...	...	...	...	...	...	...	...	...
452	Trey Lyles	Utah Jazz	41.0	PF	20.0	6-10	234.0	Kentucky	2.239800e+06
453	Shelvin Mack	Utah Jazz	8.0	PG	26.0	6-3	203.0	Butler	2.433333e+06

Next steps: Generate code with ds View recommended plots

```
ds['Height'].value_counts()
```

```
Height
6-9     59
6-10    47
6-7     45
6-8     43
6-6     42
6-11    40
6-3     33
6-5     32
6-4     29
7-0     27
6-1     16
6-2     16
6-0     10
7-1      7
7-3      4
5-11     3
7-2      3
5-9      1
Name: count, dtype: int64
```

```
height_groups = ds.groupby(ds['Height'])
height_groups
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x7828a7ab3040>
```

```
height_groups.get_group('6-11')
```

98	Jordan	Clippers	6.0	C	27.0	6-11	265.0	Texas A&M
113	Ryan Kelly	Los Angeles Lakers	4.0	PF	25.0	6-11	230.0	Duke
143	DeMarcus Cousins	Sacramento Kings	15.0	C	25.0	6-11	270.0	Kentucky
162	Joakim Noah	Chicago Bulls	13.0	C	31.0	6-11	232.0	Florida
163	Bobby Portis	Chicago Bulls	5.0	PF	21.0	6-11	230.0	Arkansas
167	Channing Frye	Cleveland Cavaliers	9.0	PF	33.0	6-11	255.0	Arizona
173	Sasha Kaun	Cleveland Cavaliers	14.0	C	31.0	6-11	260.0	Kansas
188	Andre Drummond	Detroit Pistons	0.0	C	22.0	6-11	279.0	Connecticut
204	Ian Mahinmi	Indiana Pacers	28.0	C	29.0	6-11	250.0	Kentucky
208	Myles Turner	Indiana Pacers	33.0	PF	20.0	6-11	243.0	Texas
209	Shayne Whittington	Indiana Pacers	42.0	PF	25.0	6-11	250.0	Western Michigan
211	Giannis Antetokounmpo	Milwaukee Bucks	34.0	SF	21.0	6-11	222.0	Kentucky
216	John Henson	Milwaukee Bucks	31.0	PF	25.0	6-11	229.0	North Carolina
220	Greg Monroe	Milwaukee Bucks	15.0	C	26.0	6-11	265.0	Georgetown
224	Miles Plumlee	Milwaukee Bucks	18.0	C	27.0	6-11	249.0	Duke
237	Zaza Pachulia	Dallas Mavericks	27.0	C	32.0	6-11	275.0	Kentucky
239	Dwight Powell	Dallas Mavericks	7.0	PF	24.0	6-11	240.0	Stanford
240	Charlie Villanueva	Dallas Mavericks	3.0	PF	31.0	6-11	232.0	Connecticut
251	Dwight Howard	Houston Rockets	12.0	C	30.0	6-11	265.0	Kentucky
294	LaMarcus Aldridge	San Antonio Spurs	12.0	PF	30.0	6-11	240.0	Texas
298	Tim Duncan	San Antonio Spurs	21.0	C	40.0	6-11	250.0	Wake Forest

```
height_groups.get_group('6-4')
```

	Name	Team	Number	Position	Age	Height	Weight	College	
9	Marcus Smart	Boston Celtics	36.0	PG	22.0	6-4	220.0	Oklahoma State	
17	Wayne Ellington	Brooklyn Nets	21.0	SG	28.0	6-4	200.0	North Carolina	
21	Sean Kilpatrick	Brooklyn Nets	6.0	SG	26.0	6-4	219.0	Cincinnati	
37	Jerian Grant	New York Knicks	13.0	PG	23.0	6-4	195.0	Notre Dame	
53	Kendall Marshall	Philadelphia 76ers	5.0	PG	24.0	6-4	200.0	North Carolina	
70	Norman Powell	Toronto Raptors	24.0	SG	23.0	6-4	215.0	UCLA	
103	JJ Redick	Los Angeles Clippers	4.0	SG	31.0	6-4	190.0	Duke	
104	Austin Rivers	Los Angeles Clippers	25.0	PG	23.0	6-4	200.0	Duke	
126	John Jenkins	Phoenix Suns	23.0	SG	25.0	6-4	215.0	Vanderbilt	
161	E'Twaun Moore	Chicago Bulls	55.0	SG	27.0	6-4	191.0	Purdue	
166	Matthew Dellavedova	Cleveland Cavaliers	8.0	PG	25.0	6-4	198.0	Saint Mary's	
193	Jodie Meeks	Detroit Pistons	20.0	SG	28.0	6-4	210.0	Kentucky	
214	Jared Cunningham	Milwaukee Bucks	9.0	SG	25.0	6-4	195.0	Oregon State	
258	Tony Allen	Memphis Grizzlies	9.0	SG	34.0	6-4	213.0	Oklahoma State	
288	Eric Gordon	New Orleans Pelicans	10.0	SG	27.0	6-4	215.0	Indiana	1
290	Jrue Holiday	New Orleans Pelicans	11.0	PG	25.0	6-4	205.0	UCLA	1
311	Kirk Hinrich	Atlanta Hawks	12.0	SG	35.0	6-4	190.0	Kansas	
325	Troy Daniels	Charlotte Hornets	30.0	SG	24.0	6-4	205.0	Virginia Commonwealth	
	Tyler								

```
height_groups['Salary'].describe()
```

Height	count	mean	std	min	25%	50%	75
5-11	3.0	5.891553e+05	7.926627e+05	55722.0	133733.0	211744.0	8.558720e+05
5-9	1.0	6.912869e+06	NaN	6912869.0	6912869.0	6912869.0	6.912869e+06
6-0	10.0	5.784075e+06	6.337144e+06	947276.0	2437500.0	3934473.5	4.846419e+06
6-1	16.0	5.217919e+06	4.286013e+06	700902.0	1646160.0	3402626.5	8.633373e+06
6-10	47.0	5.185375e+06	5.063120e+06	222888.0	1054584.5	3815000.0	7.025766e+06
6-11	40.0	6.544397e+06	6.906416e+06	245177.0	1362370.0	3107656.0	1.143804e+07
6-2	16.0	3.523777e+06	3.631376e+06	525093.0	947276.0	1553220.0	4.882013e+06
6-3	33.0	5.821784e+06	5.668225e+06	189455.0	1662360.0	4053446.0	8.000000e+06
6-4	29.0	4.646163e+06	5.275308e+06	134215.0	1015421.0	2525160.0	5.192520e+06
6-5	32.0	4.391786e+06	4.114296e+06	55722.0	1160040.0	3129420.0	6.015152e+06
6-6	42.0	3.586813e+06	4.518975e+06	167406.0	955794.0	1903380.0	4.317674e+06
6-7	45.0	3.504402e+06	4.337857e+06	30888.0	947276.0	1535880.0	4.000000e+06
6-8	43.0	5.950412e+06	6.133934e+06	83397.0	1259700.0	3425510.0	9.321234e+06
6-9	59.0	4.157787e+06	4.517154e+06	111444.0	1053814.0	2500000.0	5.250000e+06
7-0	27.0	5.287712e+06	4.675298e+06	947276.0	2003580.0	4204200.0	7.574380e+06
7-1	7.0	7.400988e+06	6.587462e+06	1175880.0	3441500.0	4950000.0	9.555017e+06
7-2	3.0	6.835639e+06	7.825718e+06	525093.0	2457350.0	4389607.0	9.990912e+06
7-3	4.0	2.307930e+06	1.484918e+06	1000000.0	1150000.0	2050000.0	3.207930e+06

```
age_groups = ds.groupby(ds['Age'])
age_groups
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x7828a7a0cb50>
```

```
age_groups['Salary'].describe()
```