Introduction

World of Warcraft is a massively multiplayer online role-playing game. Within this game is a sub-game type called Battlegrounds. Battlegrounds are player vs player scenarios where two teams battle each other. This dataset is a recording of many such games from one person, and we will use EDA to accomplish the goals of the project.
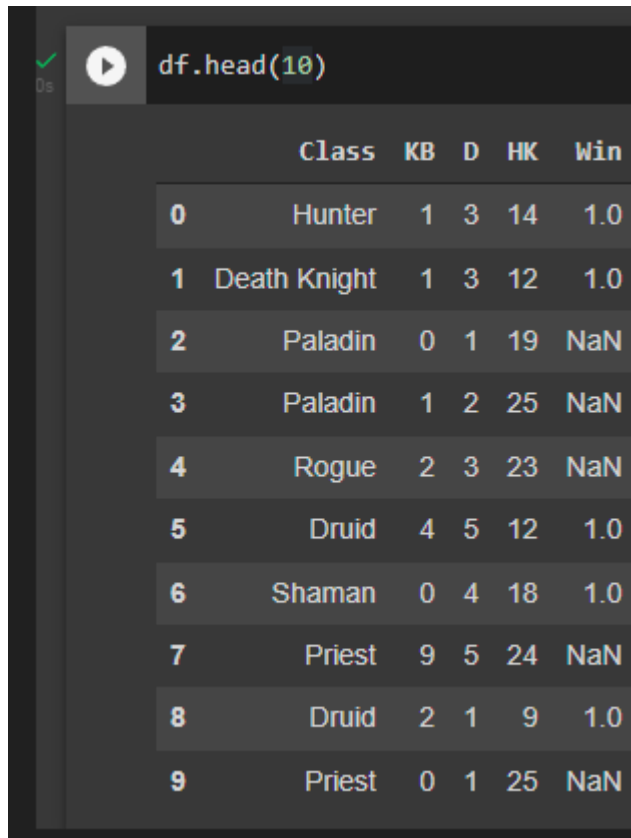
Project Goals, Plan, and Methodology

I am interested in finding the best class to play - the class that wins the most. I am secondarily interested in finding out what variables and their relationships are positively correlated to a class winning more often.

My plan is to use exploratory data analysis techniques like univariate analysis, bivariate analysis, distribution analysis, measures of center and spread of values, inference and parameter estimation, correlation analysis, and visualization. We will step through this dataset and hopefully gleam important information we can use to decide which class is the best at Battlegrounds.

Exploratory Data Analysis

The Dataset and its Variables

The first 10 row of the dataset

```
df.head(10)
```

	Class	KB	D	HK	Win
0	Hunter	1	3	14	1.0
1	Death Knight	1	3	12	1.0
2	Paladin	0	1	19	NaN
3	Paladin	1	2	25	NaN
4	Rogue	2	3	23	NaN
5	Druid	4	5	12	1.0
6	Shaman	0	4	18	1.0
7	Priest	9	5	24	NaN
8	Druid	2	1	9	1.0
9	Priest	0	1	25	NaN

Variable Description and Range Table

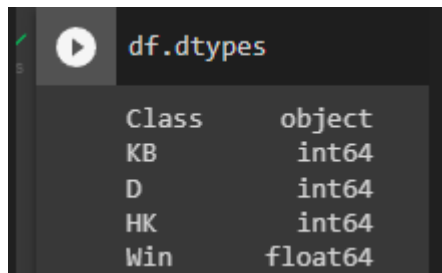| Variables	Description	Range
Class	Class of the played character	Warrior, paladin, hunter, rogue, priest, death knight, shaman, mage, warlock, monk, druid, demon hunter.
Killing Blows (KB)	Number of mortal kills this game	0-18
Deaths (D)	Number of times the player died this game	0-11
Honorable Kills (HK)	Number of mortal kills given by the player	0-68
Win	1 if the players team won	0, 1

Variable Data Types

Class has 13 defined categories - Nominal Data type
KB - Discrete integer.

D - Discrete integer.

HK - Discrete integer.

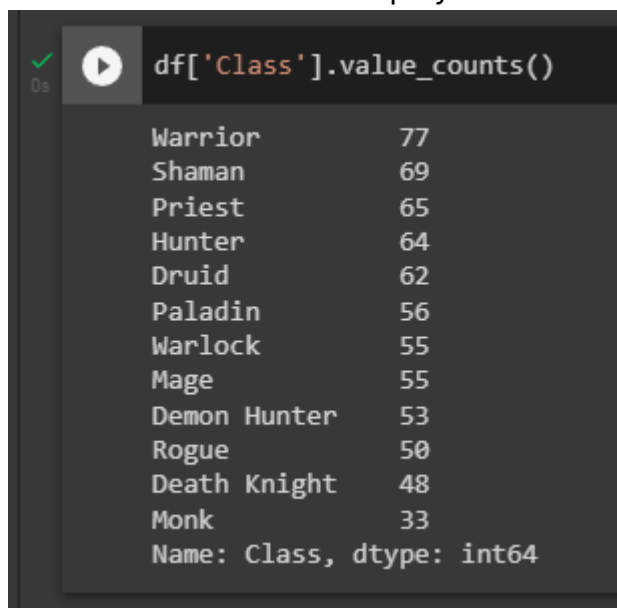Win - represented by a 1 if true, empty cell if not true.

```
df.dtypes

Class     object
KB         int64
D          int64
HK         int64
Win      float64
```

Dataset Shape

This data set has 5 columns(variables) and 687 rows(games played)

Class

There are 12 playable classes in World of Warcraft. Warrior is this players most played class while Monk is least played.

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

Warrior          77
Shaman           69
Priest           65
Hunter           64
Druid            62
Paladin          56
Warlock          55
Mage             55
Demon Hunter     53
Rogue            50
Death Knight     48
Monk             33
Name: Class, dtype: int64
```

Killing Blows

To my logic this variable should be positively correlated to our win rate - the more enemies the player kills, the better the chance of winning.
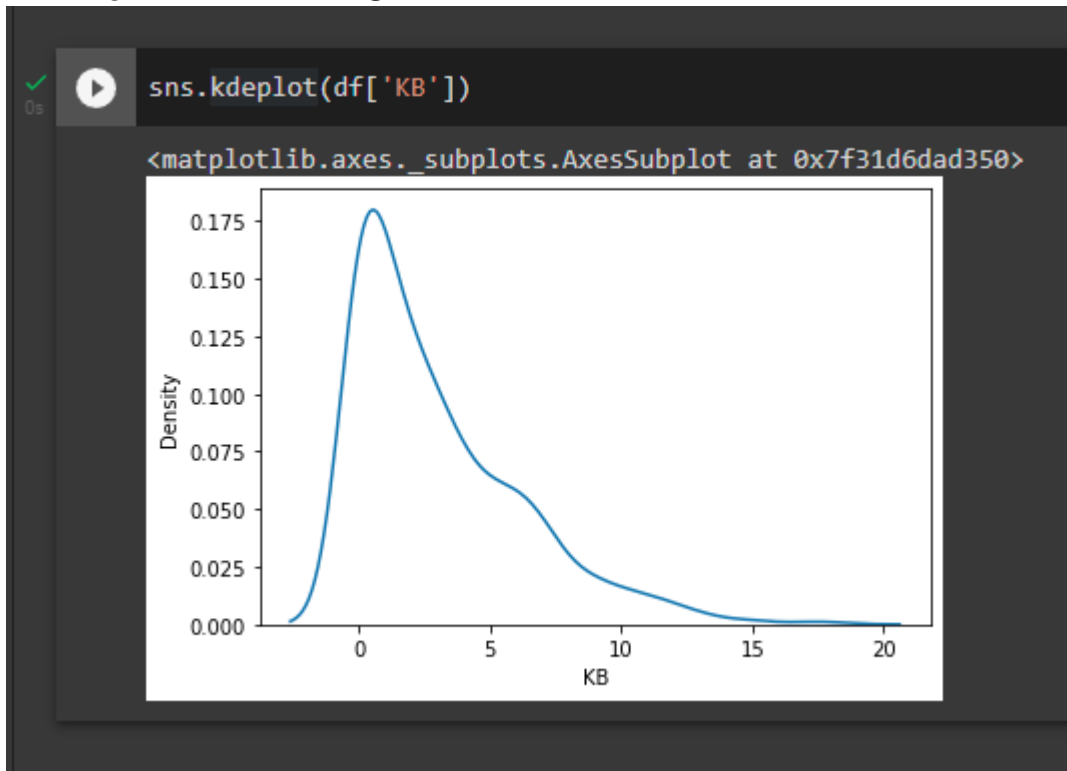
Median, Mean and Standard Deviation

```
[15] df['KB'].median()

     2.0
```

```
[16] df['KB'].mean()

     3.039301310043668
```

```
     df['KB'].std()

     3.2175581674125833
```

Frequency Table for Killing Blows

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

     0      179
     1      115
     2       85
     3       73
     4       50
     6       43
     5       41
     7       35
     8       16
     9       15
     10      10
     11      10
     12       6
     13       4
     15       2
     17       1
     14       1
     18       1
     Name: KB, dtype: int64
```
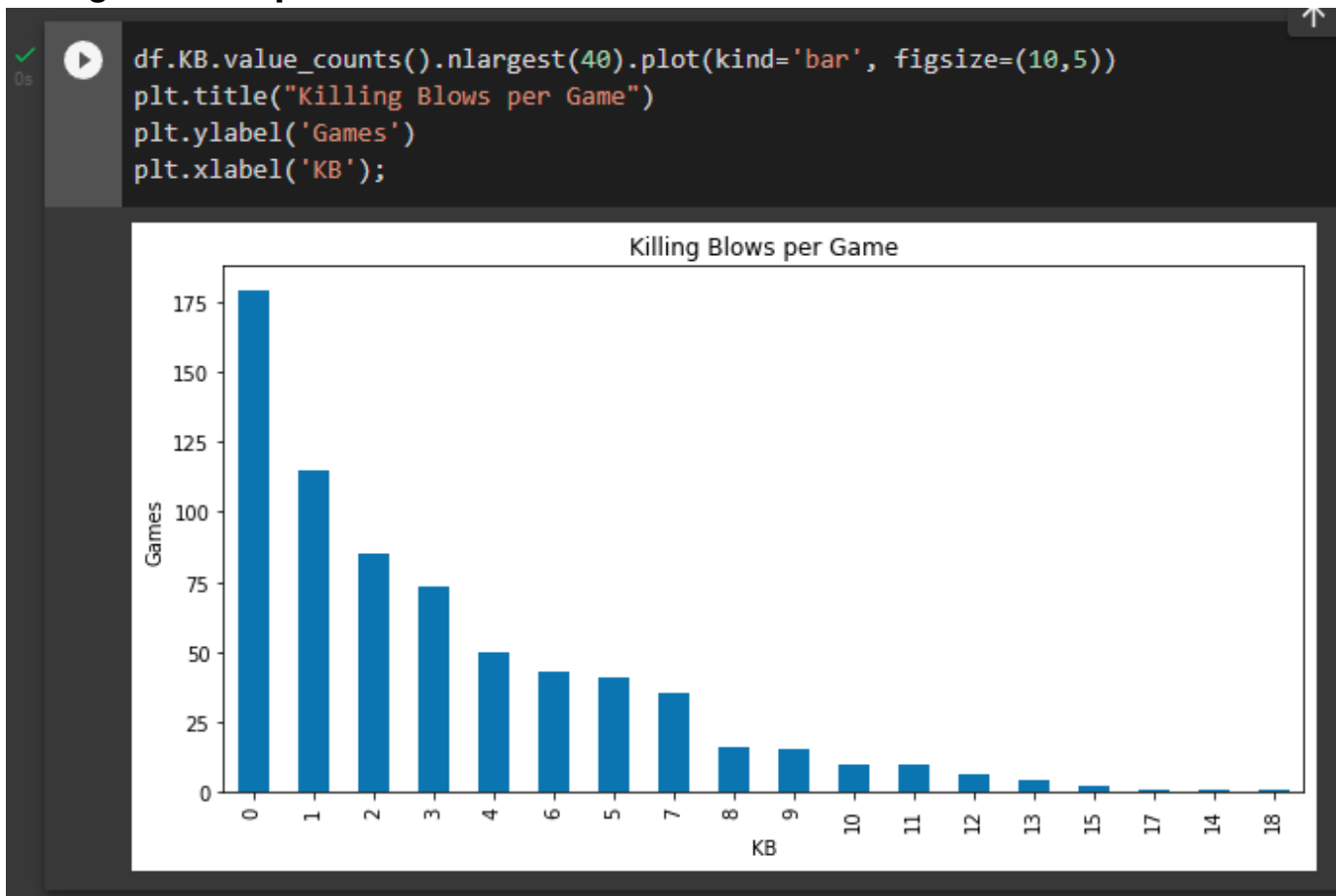
Density curve for Killing Blows variable

```
sns.kdeplot(df['KB'])
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f31d6dad350>

Histogram for KB per Game

```
df.KB.value_counts().nlargest(40).plot(kind='bar', figsize=(10,5))
plt.title("Killing Blows per Game")
plt.ylabel('Games')
plt.xlabel('KB');
```


Deaths(D)

To my logic this variable should be negatively correlated to our win rate - the more the player dies, the less chance of winning.

Median, Mean and Standard Deviation

```
[23] df['D'].median()

     3.0

[24] df['D'].mean()

     2.997088791848617

[25] df['D'].std()

     2.304133115379255
```
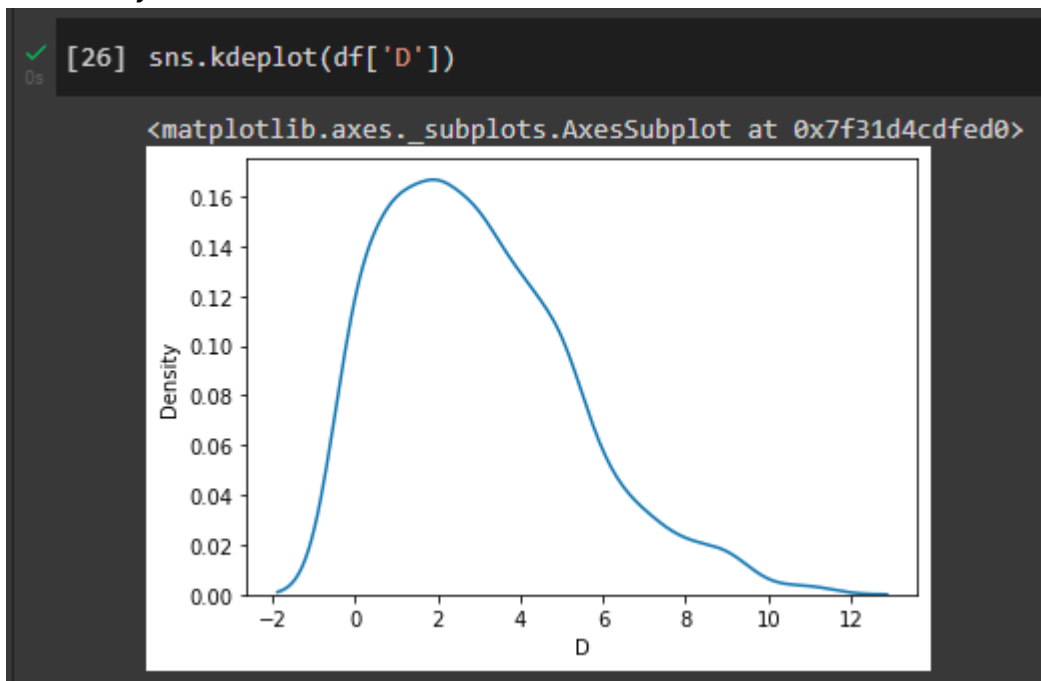
**Frequency Table

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

     2      117
     1      112
     3      108
     0       96
     4       87
     5       77
     6       34
     7       23
     9       14
     8       14
     11       3
     10       2
     Name: D, dtype: int64
```

**Density curve

```
[26] sns.kdeplot(df['D'])
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f31d4cdfed0>

Honorable Kills(HK)w

Median, Mean and Standard Deviation

```
[28] df['HK'].median()
```
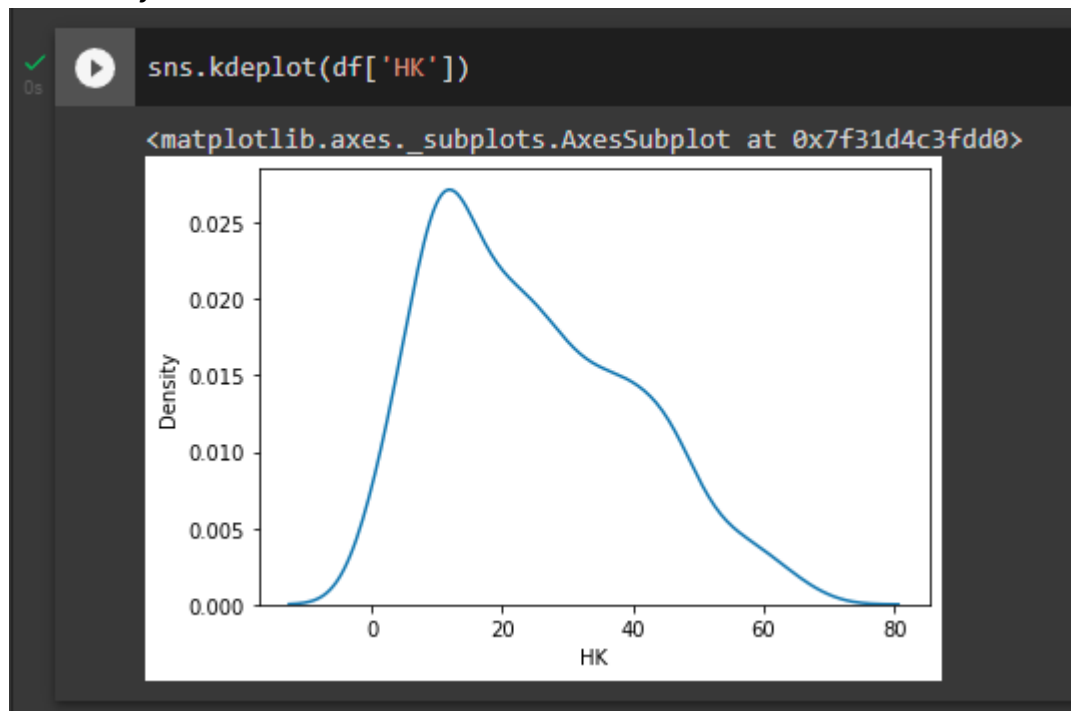
22.0

```
[29] df['HK'].mean()
```

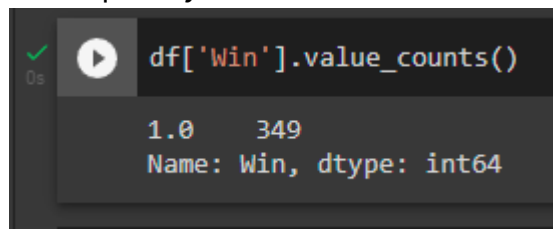24.737991266375545

```
[30] df['HK'].std()
```

15.509652709021001

**Density curve

```
sns.kdeplot(df['HK'])
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f31d4c3fdd0>

Win

The player won 349 out of 687 games, for a win rate of 50.8%.

**Frequency Table

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

```
1.0    349
Name: Win, dtype: int64
```
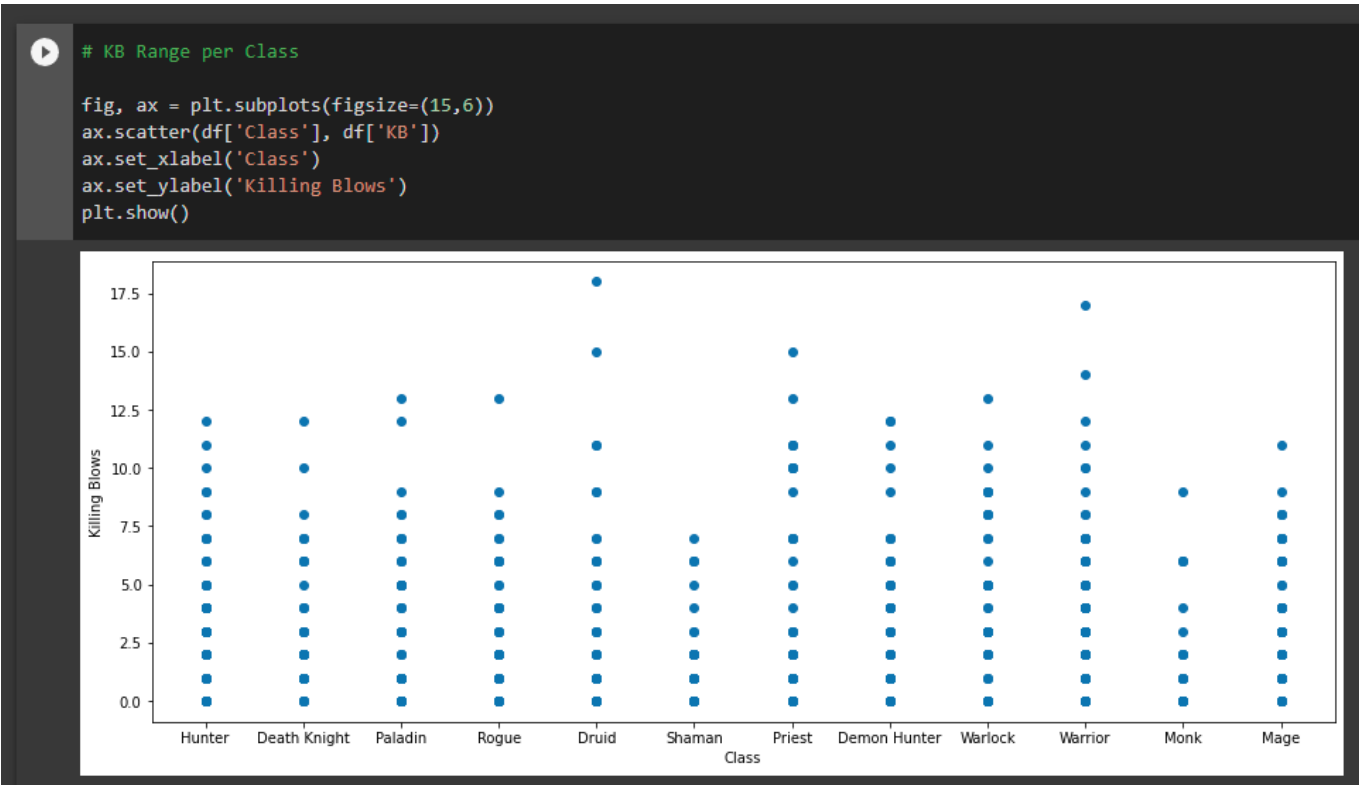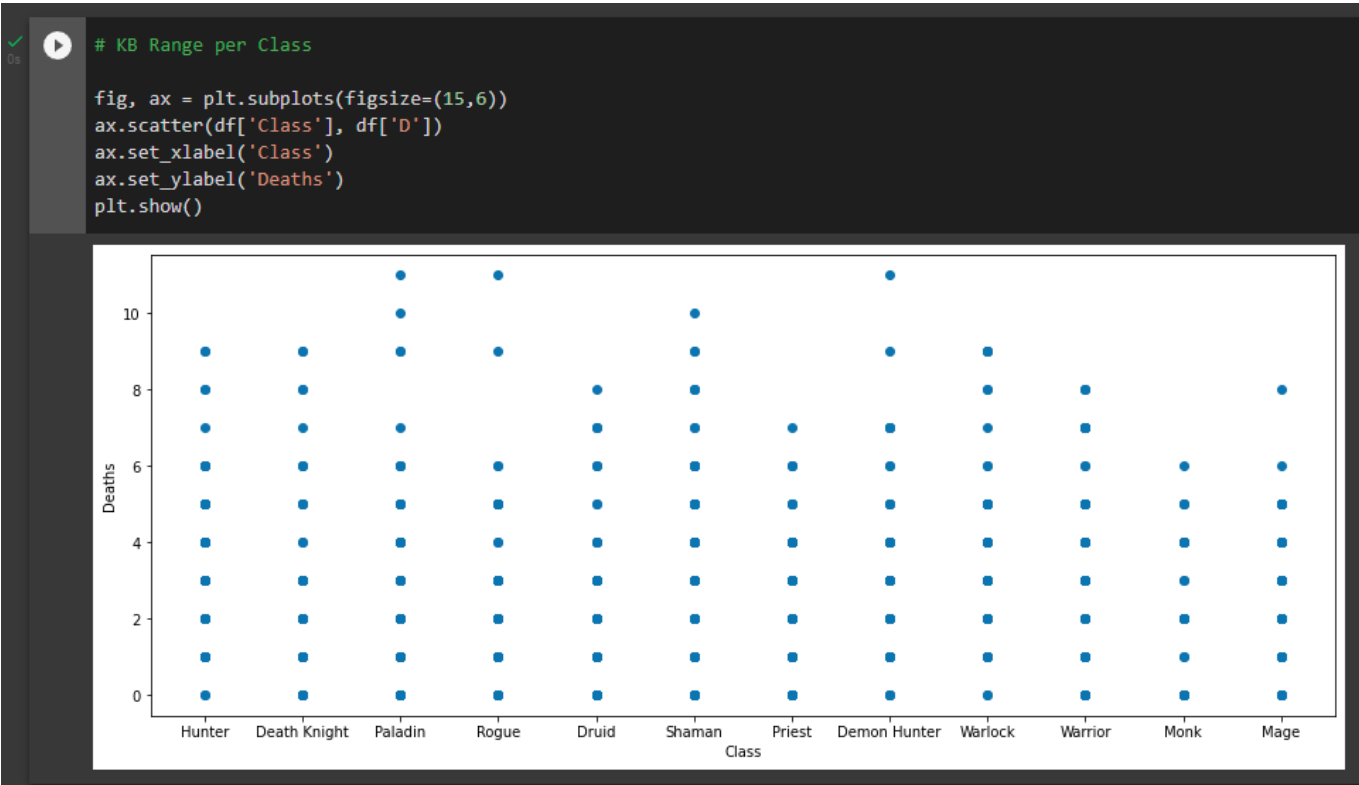
Bivariate Analysis

Now that we have a basic sense of the single variables, lets use bivariate analysis to explore the relationships of the variable pairs I feel will be more likely to effect each Class' Win Percentage.

Scatterplot of Killing Blow Range per Class

```
# KB Range per Class

fig, ax = plt.subplots(figsize=(15,6))
ax.scatter(df['Class'], df['KB'])
ax.set_xlabel('Class')
ax.set_ylabel('Killing Blows')
plt.show()
```


Scatterplot of Death Range per Class

```
# KB Range per Class

fig, ax = plt.subplots(figsize=(15,6))
ax.scatter(df['Class'], df['D'])
ax.set_xlabel('Class')
ax.set_ylabel('Deaths')
plt.show()
```

Scatterplot of Honorable Kill Range per Class

```
# KB Range per Class

fig, ax = plt.subplots(figsize=(15,6))
ax.scatter(df['Class'], df['HK'])
ax.set_xlabel('Class')
ax.set_ylabel('Honorable Kills')
plt.show()
```
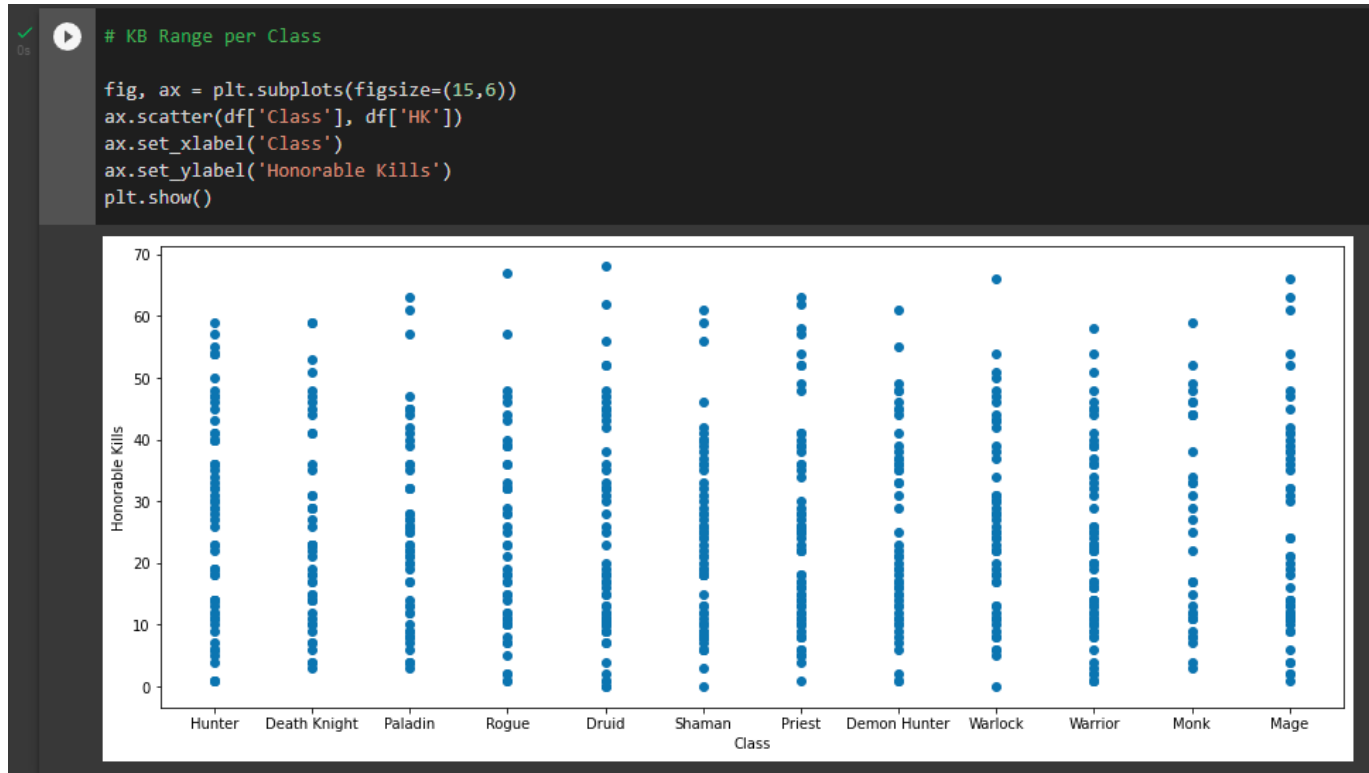

Class	Mean Killing Blows(KB)	Mean Deaths(D)	KB/D Ratio	Win %
Warrior	4.0909	3.4156	1.1977	46.76%
Paladin	2.8929	3.0000	0.9643	51.79%
Hunter	3.5625	3.5625	1.0000	48.43%
Rogue	2.9800	2.74	1.0880	54.00%
Priest	2.9692	2.3692	1.2533	53.85%
Death Knight	2.9375	2.9375	1.0000	56.25%
Shaman	1.1159	3.4058	0.3276	53.62%
Mage	3.5454	2.6727	1.3265	40.00%
Warlock	4.0727	3.7091	1.0980	40.00%
Monk	1.4242	2.0303	0.7015	54.54%
Druid	2.7581	2.3710	1.1633	53.23%
Demon Hunter	3.5094	3.1698	1.107	60.04%

Our top three classes by win rate are Demon Hunter, Death Knight, then Monk.

Summary and Recommendations based on EDA

Based on our analysis Killing Blows, Honorable Kills, and Deaths are integer variables that possibly could have been correlated with each class's win rate.

The Class with the highest KB average was Warrior, a fairly low Win % class. Second was Warlock, an even lower Win % class at 40%. There does not seem to be any strong correlation between KB and Win%.

The Classes with the highest HK game was the Druid(53.23% Win), the Rogue was next(54% Win).

The Class with the lowest Mean Deaths(D) was Monk, who has a fairly high Win % of 54.54%, Next Lowest was Priest, also a fairly high Win %. The highest D rate went to the Warlock with 3.7091 who also had the tied lowest Win %. There does seems to be some correlation between Mean Deaths and Win %.

Interestingly, the player's most played class, the warrior, has the third lowest win rate of 46.76%

The Highest Win % Class was the Demon Hunter, and based on this EDA I would consider it the best class to play battlegrounds. A class is generally strong if it has a low death rate.