RS3 Final Project

Analyzing Effect of Fixture Congestion on Premier League and Whether Or Not It Impacts Fatigue Based On Defensive Stats

Saurav Banerjee

Joshua Park

In [ ]:

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import numpy as np
import matplotlib.pyplot as plt
import math
import scipy.stats
%matplotlib inline
```

In [ ]:

```
data1718 = pd.read_csv('1718processed.csv')
data1819 = pd.read_csv('1819processed.csv')
data1920 = pd.read_csv('1920processed.csv')
```

In [ ]:

```
data1718
```

Out[ ]:

	Squad	#Pl	TacklesTkl	TacklesTklW	TacklesDef3rd	TacklesMid3rd	TacklesAtt3rd	VsDribblesTkl	VsDribblesAtt	VsDribblesTkl%	VsD
0	Arsenal	30	690	436	326	268	96	262	667	39.3	
1	Bournemouth	22	582	331	298	212	72	191	551	34.7	
2	Brighton	24	692	447	411	235	46	230	727	31.6	
3	Burnley	24	578	351	297	223	58	220	632	34.8	
4	Chelsea	26	720	445	357	285	78	238	583	40.8	
5	Crystal Palace	28	760	451	429	268	63	258	682	37.8	
6	Everton	30	749	459	408	260	81	247	672	36.8	
7	Huddersfield	25	829	464	388	349	92	241	674	35.8	
8	Leicester City	27	703	433	376	261	66	203	662	30.7	

	Squad	#Pl	TacklesTkl	TacklesTklW	TacklesDef3rd	TacklesMid3rd	TacklesAtt3rd	VsDribblesTkl	VsDribblesAtt	VsDribblesTkl%	VsD
9	Liverpool	27	706	448	299	301	106	207	564	36.7	
10	Manchester City	25	641	406	259	275	107	219	609	36.0	
11	Manchester Utd	27	600	375	273	250	77	192	581	33.0	
12	Newcastle Utd	27	728	410	392	257	79	230	660	34.8	
13	Southampton	26	766	482	410	274	82	233	632	36.9	
14	Stoke City	30	768	435	423	276	69	253	643	39.3	
15	Swansea City	26	659	394	313	258	88	214	603	35.5	
16	Tottenham	25	695	422	286	295	114	225	601	37.4	
17	Watford	29	736	454	376	284	76	219	606	36.1	
18	West Brom	24	669	397	369	227	73	259	700	37.0	
19	West Ham	27	662	411	349	238	75	220	646	34.1	

In [ ]:

```
data1819
```

Out[ ]:

	Squad	#Pl	TacklesTkl	TacklesTklW	TacklesDef3rd	TacklesMid3rd	TacklesAtt3rd	VsDribblesTkl	VsDribblesAtt	VsDribblesTkl%	VsD
0	Arsenal	28	623	398	305	238	80	207	570	36.3	
1	Bournemouth	28	547	341	290	196	61	173	496	34.9	
2	Brighton	21	712	439	396	226	90	205	587	34.9	
3	Burnley	23	618	355	301	254	63	223	587	38.0	
4	Cardiff City	25	708	417	377	252	79	200	591	33.8	
5	Chelsea	24	655	426	317	251	87	204	534	38.2	
6	Crystal Palace	26	777	457	428	262	87	228	618	36.9	
7	Everton	23	777	458	389	283	105	216	568	38.0	

	Squad	#Pl	TacklesTkl	TacklesTklW	TacklesDef3rd	TacklesMid3rd	TacklesAtt3rd	VsDribblesTkl	VsDribblesAtt	VsDribblesTkl%	VsD
8	Fulham	28	649	411	356	240	53	201	537	37.4	
9	Huddersfield	31	763	430	366	310	87	224	691	32.4	
10	Leicester City	25	730	438	394	264	72	214	568	37.7	
11	Liverpool	23	617	380	265	257	95	193	571	33.8	
12	Manchester City	21	529	359	212	229	88	182	498	36.5	
13	Manchester Utd	28	619	378	319	230	70	205	542	37.8	
14	Newcastle Utd	24	736	431	398	269	69	223	640	34.8	
15	Southampton	30	753	437	374	276	103	219	648	33.8	
16	Tottenham	28	666	408	339	220	107	215	570	37.7	
17	Watford	25	689	432	365	249	75	215	617	34.8	
18	West Ham	26	748	459	406	273	69	244	729	33.5	
19	Wolves	21	750	425	395	287	68	277	721	38.4	

In [ ]:

```
data1920
```

Out[ ]:

	Squad	#Pl	TacklesTkl	TacklesTklW	TacklesDef3rd	TacklesMid3rd	TacklesAtt3rd	VsDribblesTkl	VsDribblesAtt	VsDribblesTkl%	VsD
0	Arsenal	27	560	344	257	207	96	178	410	43.4	
1	Aston Villa	30	668	390	335	254	79	244	567	43.0	
2	Brentford	29	681	391	333	248	100	254	591	43.0	
3	Brighton	26	710	444	319	274	117	237	562	42.2	
4	Burnley	23	618	354	292	247	79	220	548	40.1	
5	Chelsea	25	634	382	266	267	101	203	516	39.3	
6	Crystal Palace	24	665	411	327	252	86	220	501	43.9	

	Squad	#Pl	TacklesTkl	TacklesTklW	TacklesDef3rd	TacklesMid3rd	TacklesAtt3rd	VsDribblesTkl	VsDribblesAtt	VsDribblesTkl%	VsD
7	Everton	32	736	462	376	265	95	287	672	42.7	
8	Leeds United	29	846	466	434	313	99	367	857	42.8	
9	Leicester City	28	735	429	351	292	92	255	647	39.4	
10	Liverpool	27	575	358	206	257	112	191	538	35.5	
11	Manchester City	26	496	318	172	222	102	183	471	38.9	
12	Manchester Utd	28	619	379	326	219	74	220	541	40.7	
13	Newcastle Utd	29	684	394	368	242	74	245	605	40.5	
14	Norwich City	28	667	387	349	254	64	253	629	40.2	
15	Southampton	25	647	391	300	260	87	217	574	37.8	
16	Tottenham	25	682	367	313	283	86	229	556	41.2	
17	Watford	30	671	407	378	233	60	241	581	41.5	
18	West Ham	25	539	340	275	193	71	216	551	39.2	
19	Wolves	26	715	455	403	245	67	246	602	40.9	

In [ ]:

```
from scipy.stats import f_oneway
from scipy import mean
```

In [ ]:

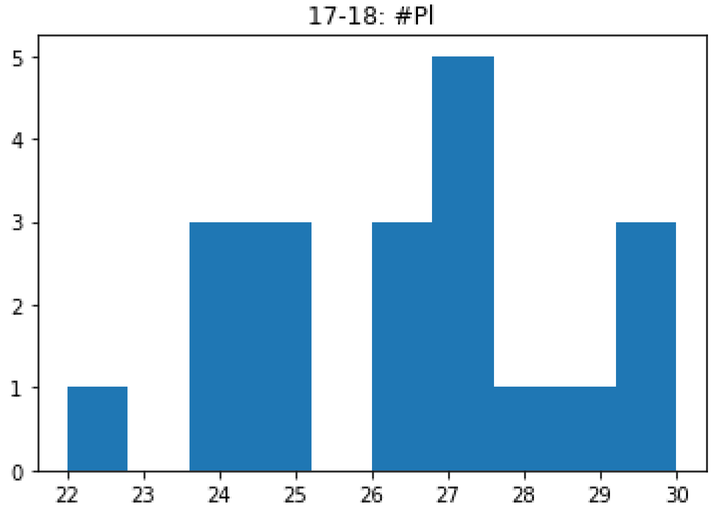
```
statnames = data1920.columns[1:]
```

In [ ]:

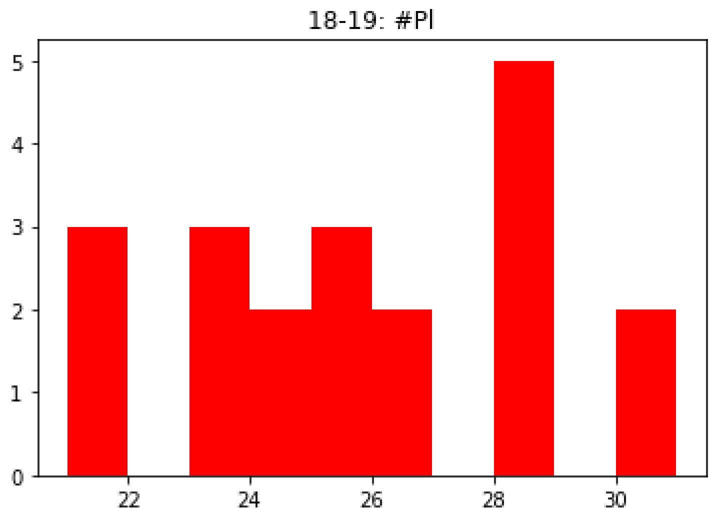
```
for x in range(len(statnames)):
    print(statnames[x])
    plt.hist(data1718.loc[:,statnames[x]])
    plt.title("17-18: " + statnames[x])
    plt.show()
    print("Mean " + str(np.mean(data1718.loc[:,statnames[x]])))
    plt.hist(data1819.loc[:,statnames[x]],color = "red")
```

```
    plt.title("18-19: " + statnames[x])
    plt.show()
    print("Mean " + str(np.mean(data1819.loc[:,statnames[x]])))
    plt.hist(data1920.loc[:,statnames[x]],color = "green")
    plt.title("19-20: " + statnames[x])
    plt.show()
    print("Mean " + str(np.mean(data1920.loc[:,statnames[x]])))
    print("ANOVA Values")
    print(f_oneway(data1718.loc[:,statnames[x]], data1819.loc[:,statnames[x]],data1819.loc[:,statnames[x]]))
```

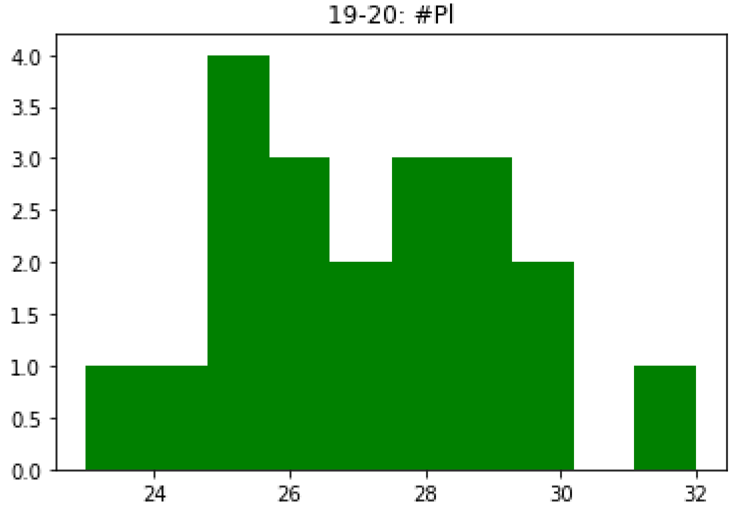
#Pl



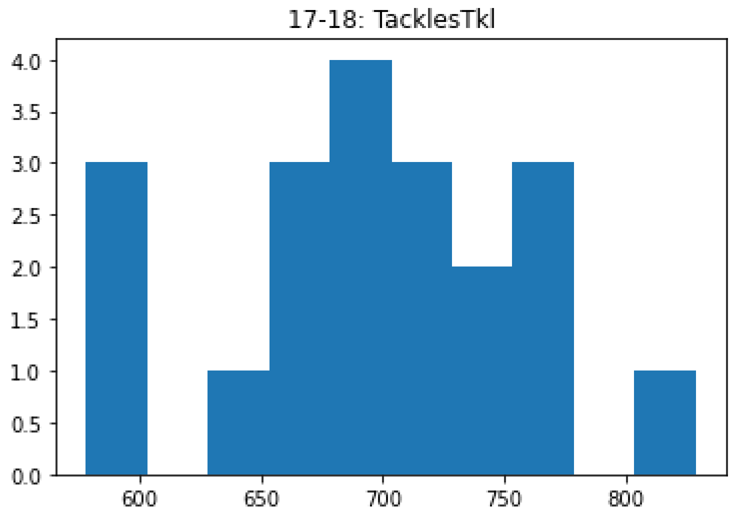
Mean 26.45



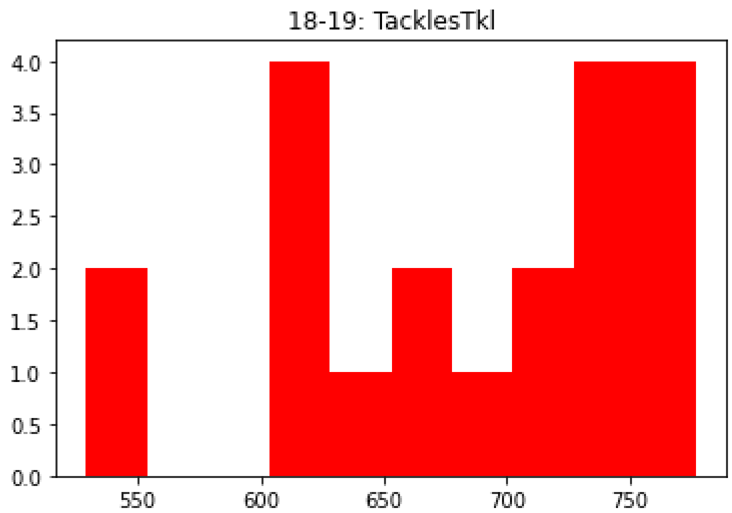
Mean 25.4



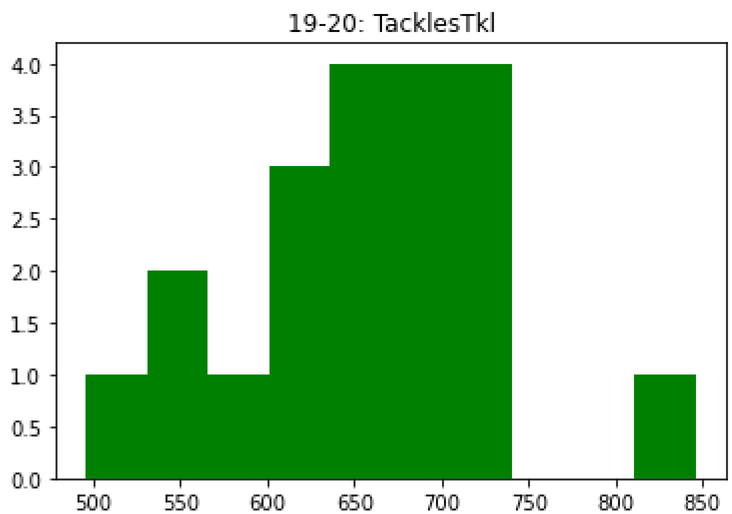
Mean 27.1
ANOVA Values
F_onewayResult(statistic=0.9821826280623608, pvalue=0.3807404291169192)
TacklesTkl



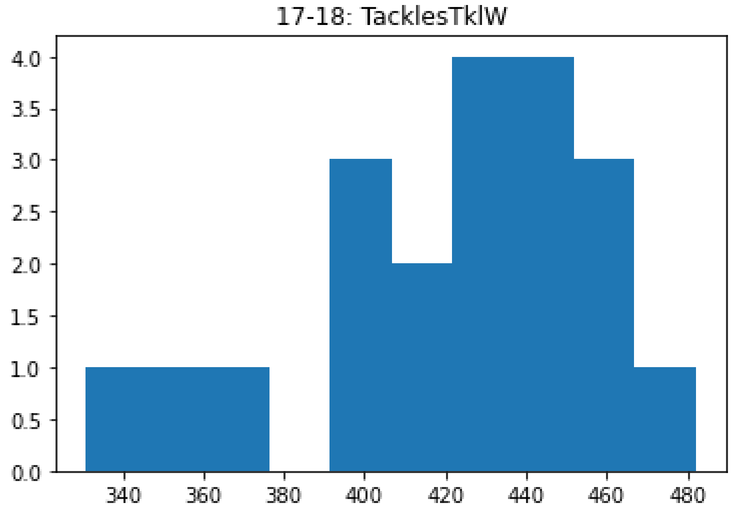
Mean 696.65



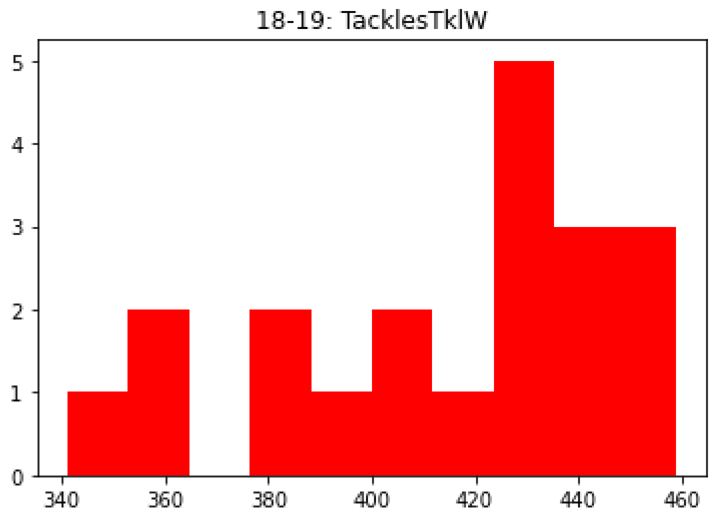
Mean 683.3



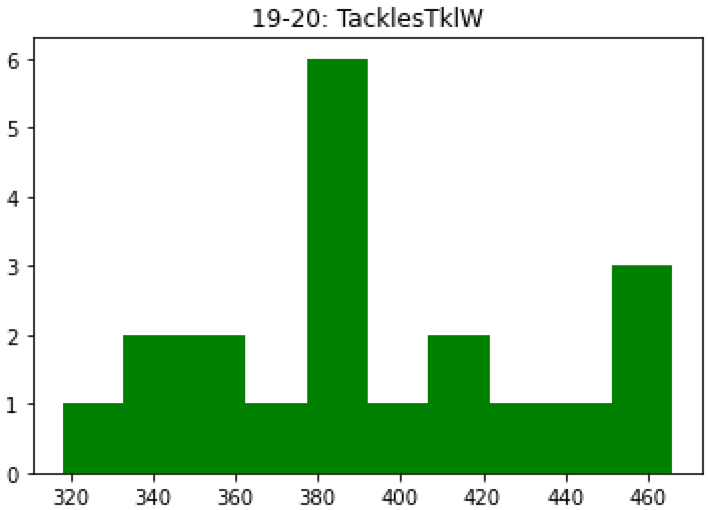
Mean 657.4
ANOVA Values
F_onewayResult(statistic=0.2332667254386761, pvalue=0.7926945755671473)
TacklesTklW



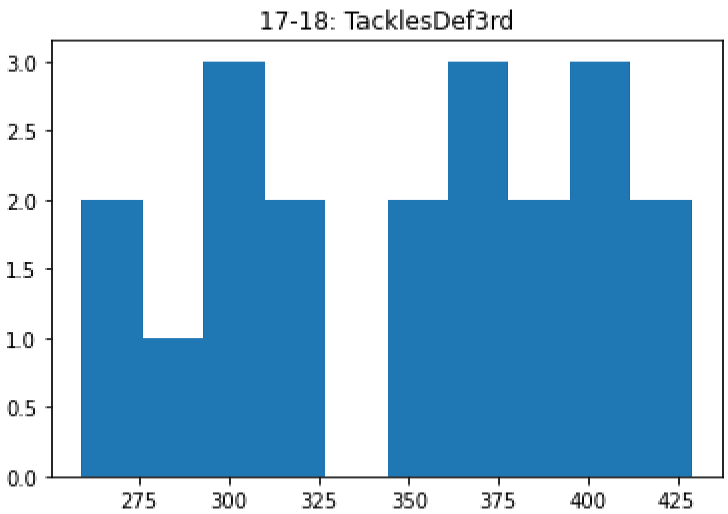
Mean 422.55



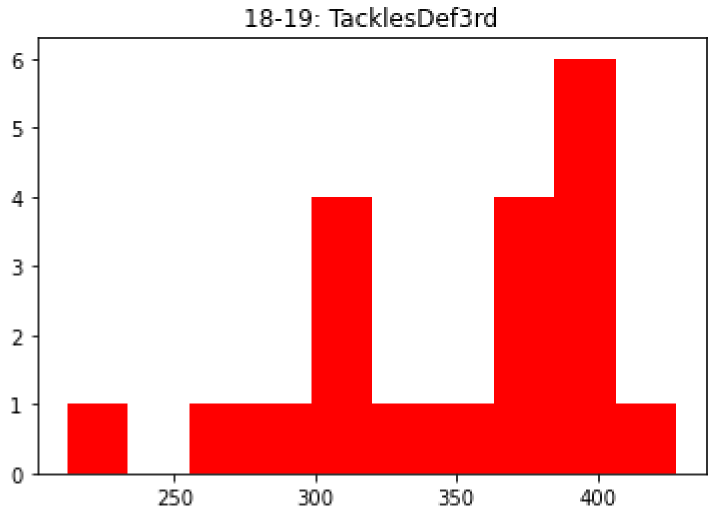
Mean 413.95



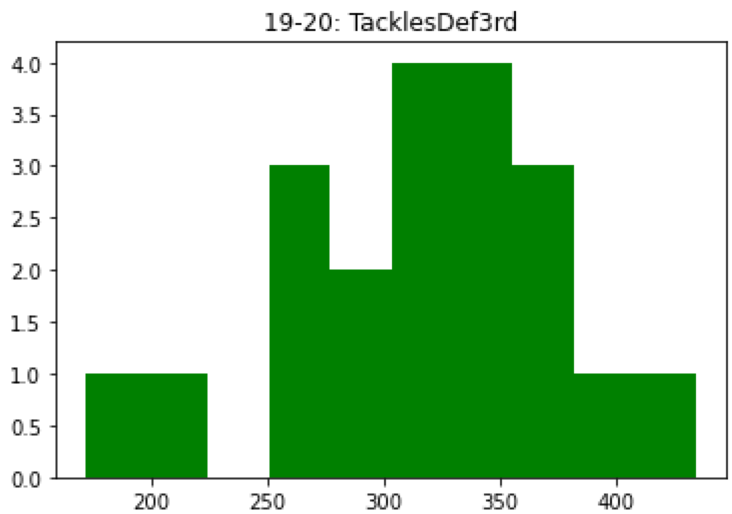

```
Mean 393.45
ANOVA Values
F_onewayResult(statistic=0.3751766266168386, pvalue=0.6888521227568817)
TacklesDef3rd
```



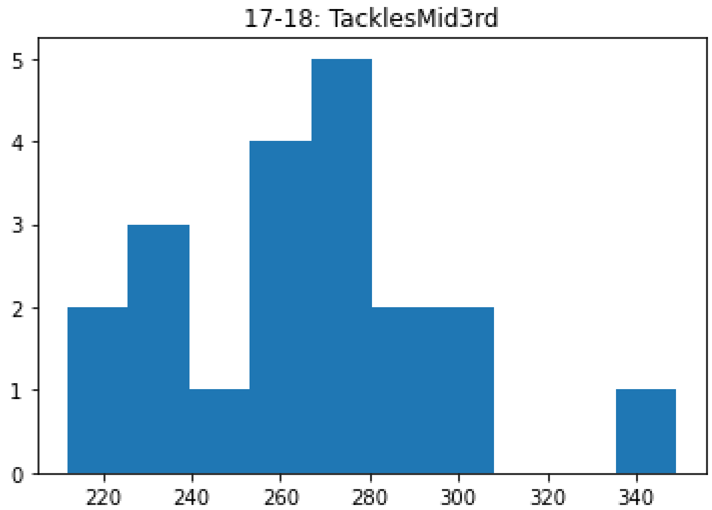

```
Mean 351.95
```



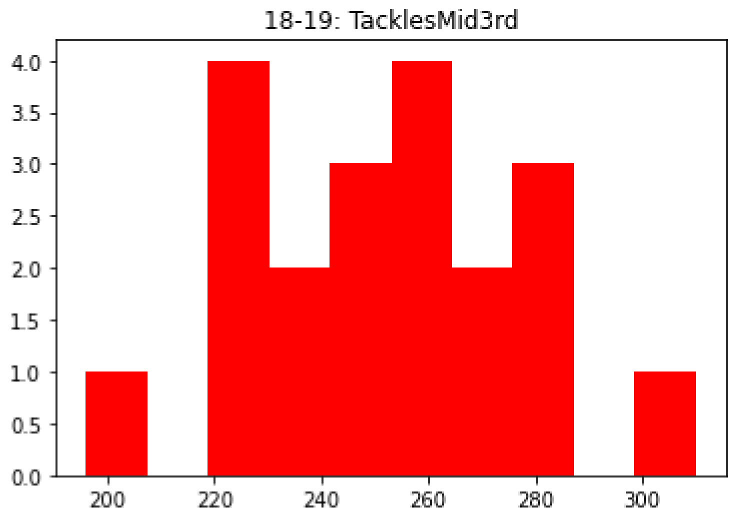

```
Mean 349.6
```



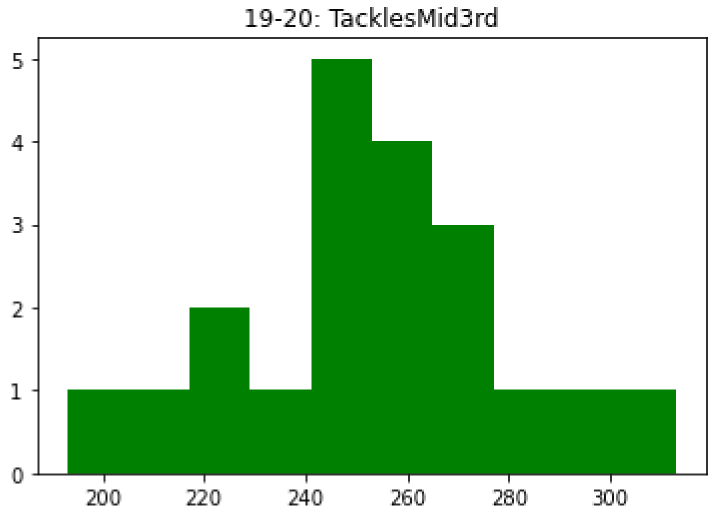

```
Mean 319.0
ANOVA Values
F_onewayResult(statistic=0.012320600392594194, pvalue=0.9877576172012333)
TacklesMid3rd
```



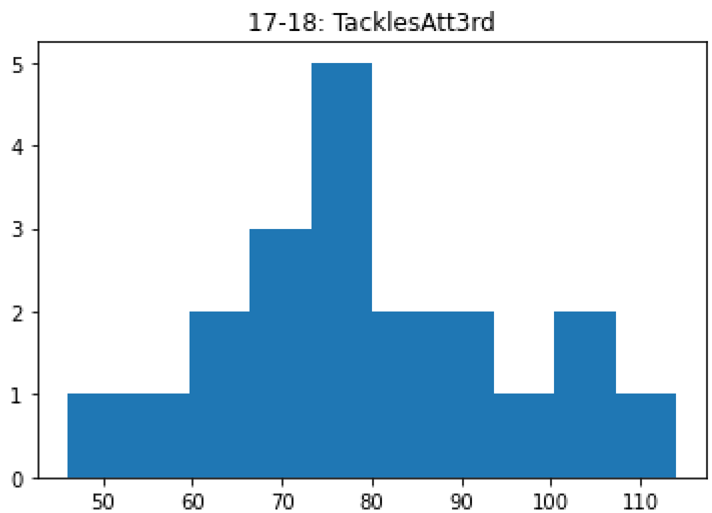
Mean 264.8



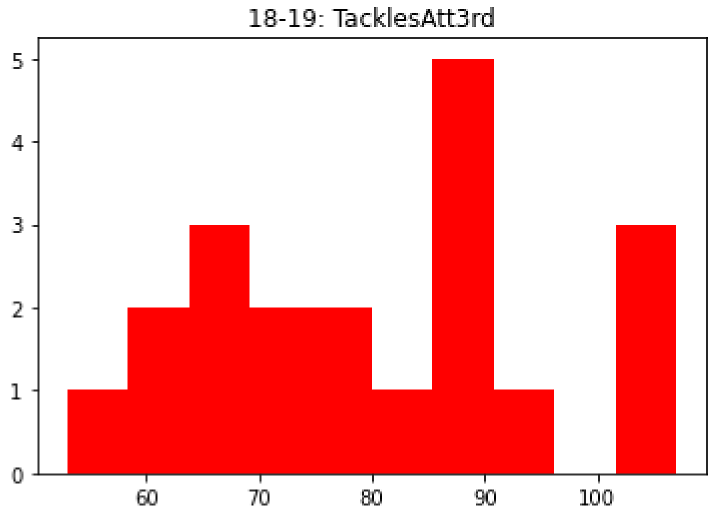
Mean 253.3



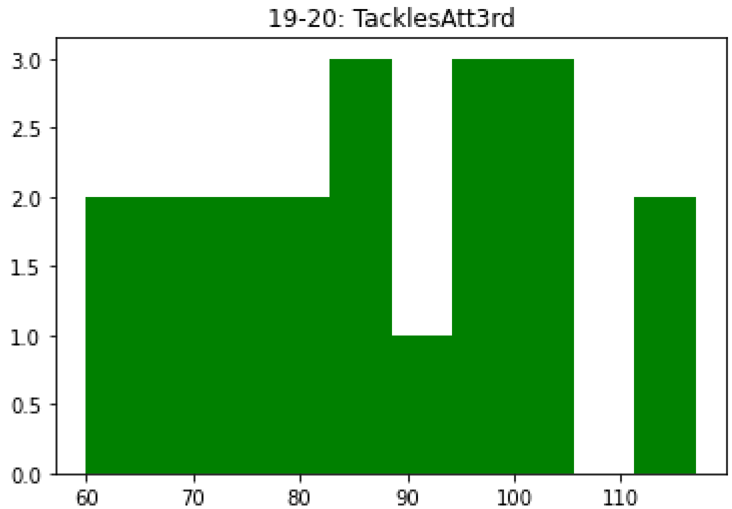
Mean 251.35
ANOVA Values
F_onewayResult(statistic=1.118129472072906, pvalue=0.3339536030136585)
TacklesAtt3rd



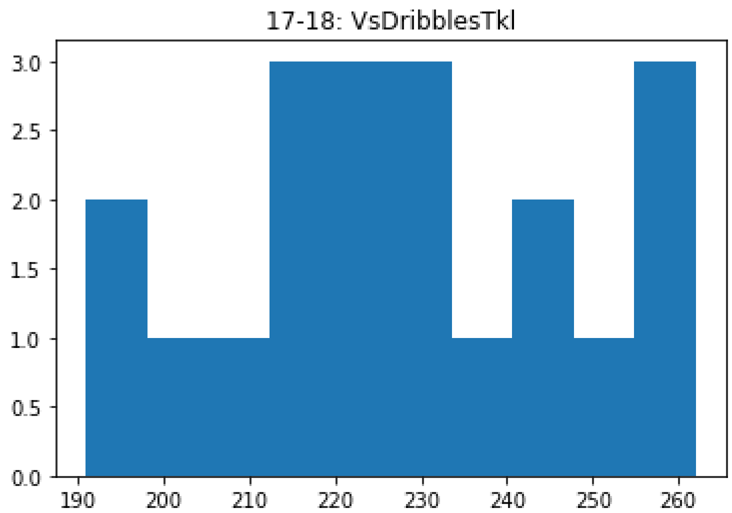
Mean 79.9



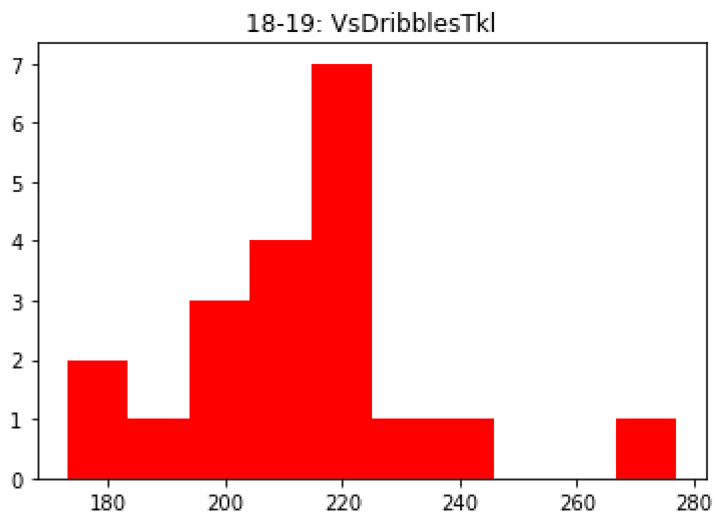
Mean 80.4



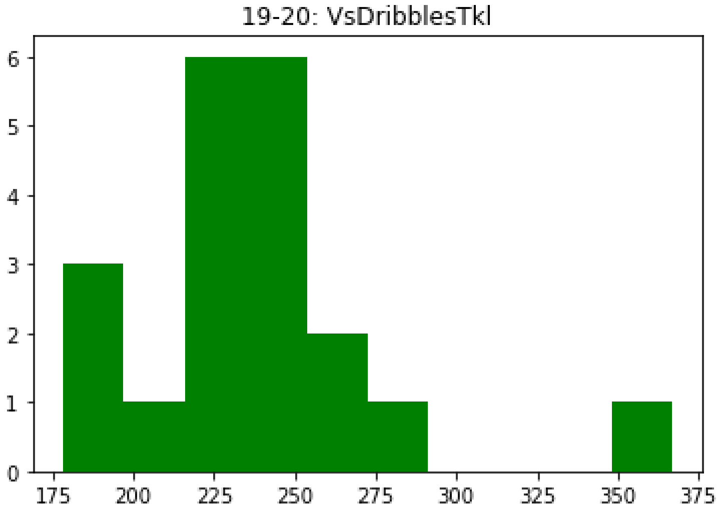

```
Mean 87.05
ANOVA Values
F_onewayResult(statistic=0.006697026520225019, pvalue=0.9933261300560847)
VsDribblesTkl
```



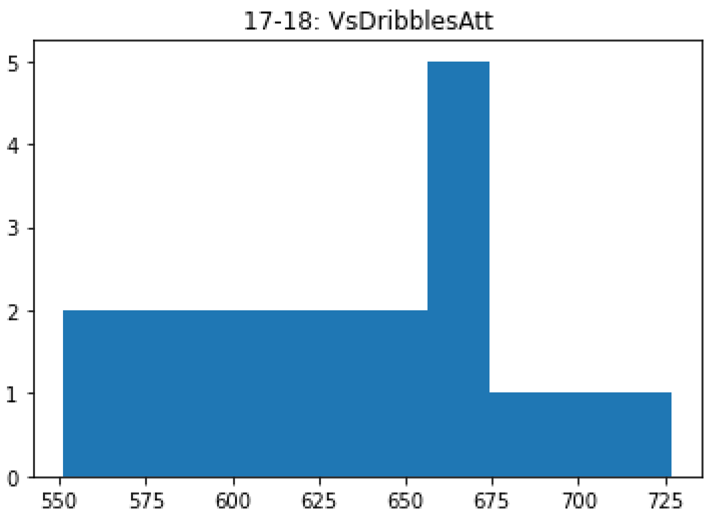

```
Mean 228.05
```



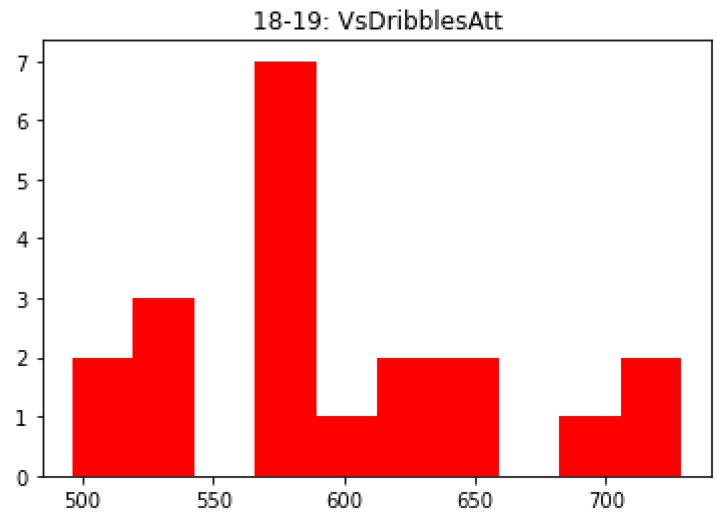

```
Mean 213.4
```



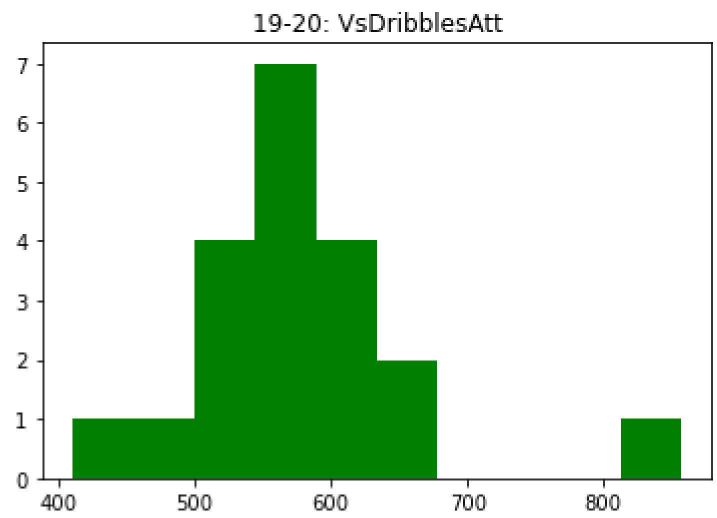

```
Mean 235.3
ANOVA Values
F_onewayResult(statistic=3.0344873132898385, pvalue=0.05593503742377694)
VsDribblesAtt
```



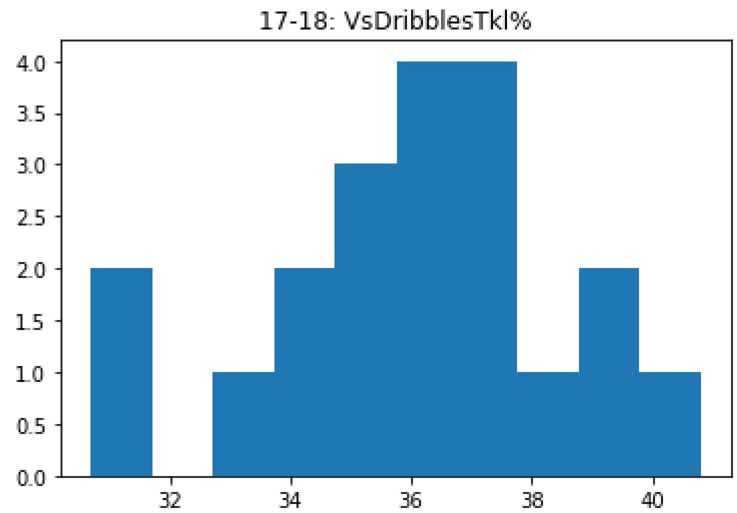

```
Mean 634.75
```



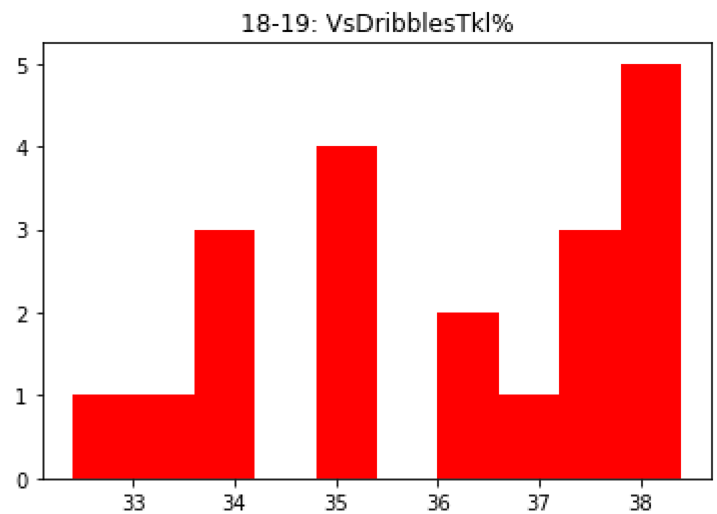
Mean 594.15



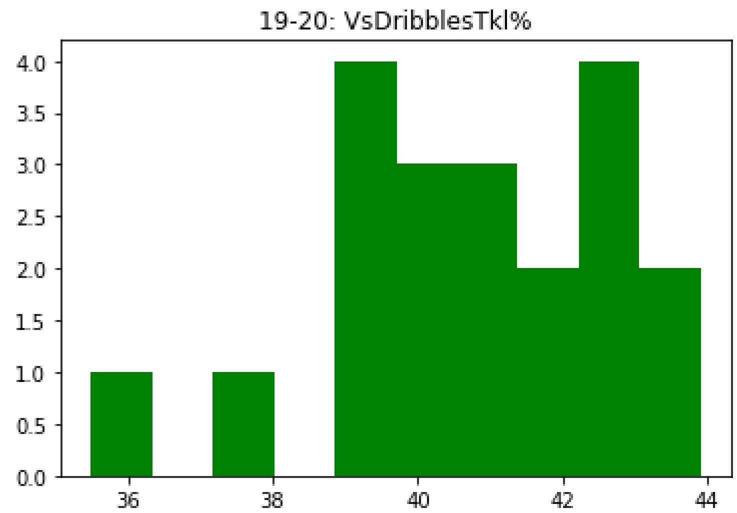
Mean 575.95
ANOVA Values
F_onewayResult(statistic=3.061367691257648, pvalue=0.054593104048497404)
VsDribblesTkl%



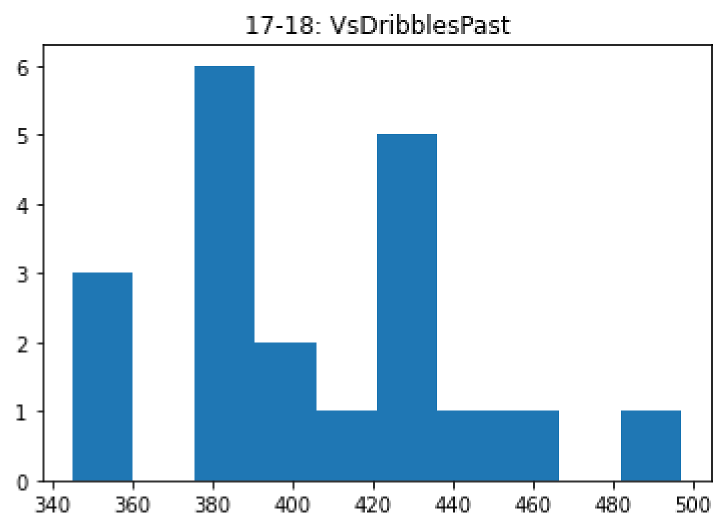
Mean 35.955



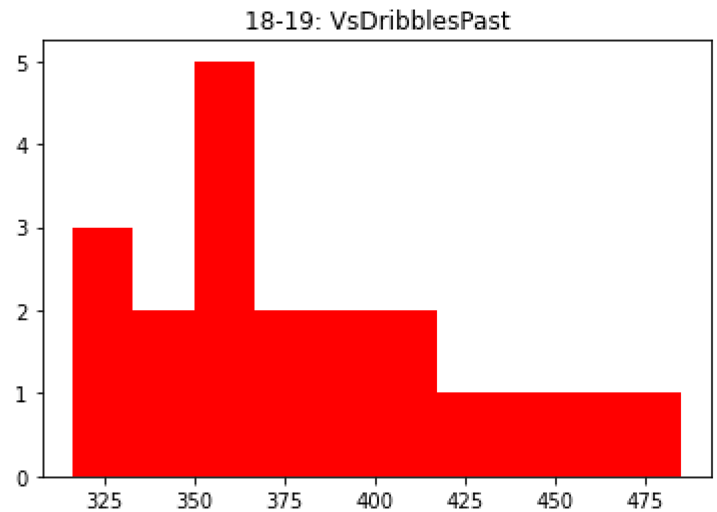
Mean 35.98



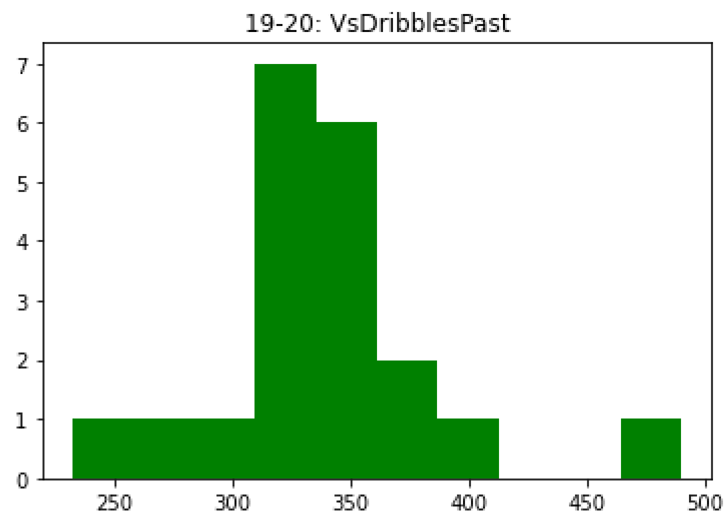
Mean 40.81
ANOVA Values
F_onewayResult(statistic=0.0009228551389431133, pvalue=0.9990775854882242)
VsDribblesPast



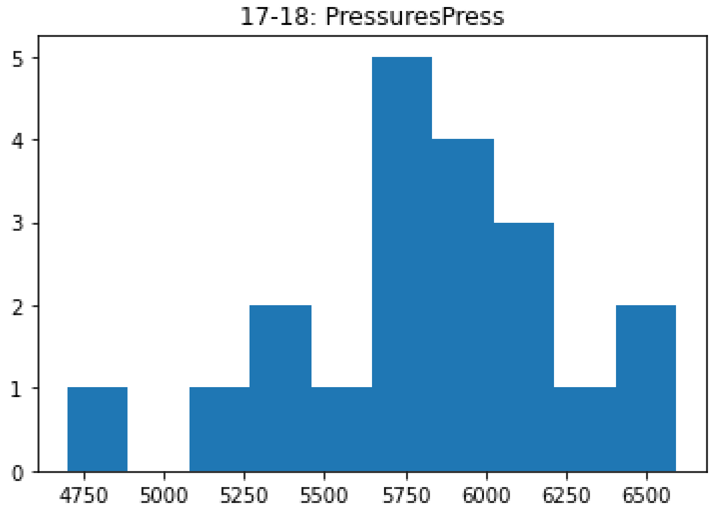
Mean 406.7



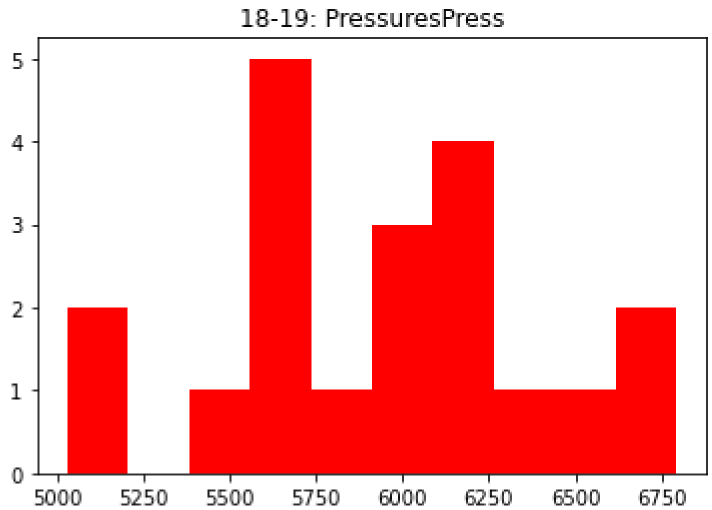
Mean 380.75



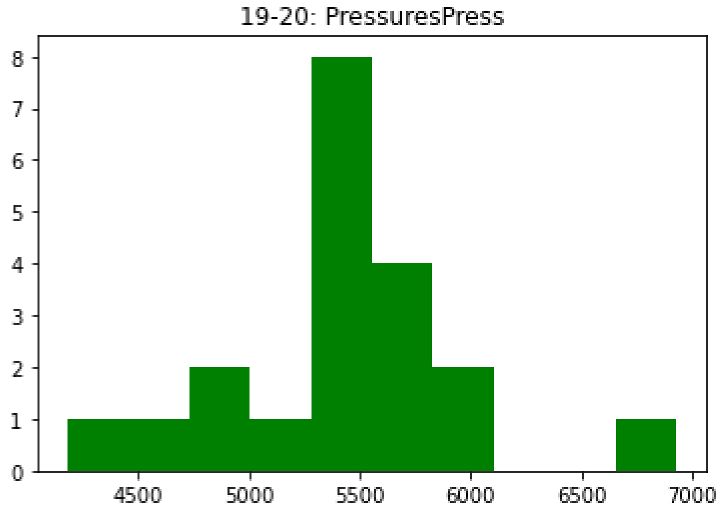
Mean 340.65
ANOVA Values
F_onewayResult(statistic=2.280407383412186, pvalue=0.111495965036905)
PressuresPress



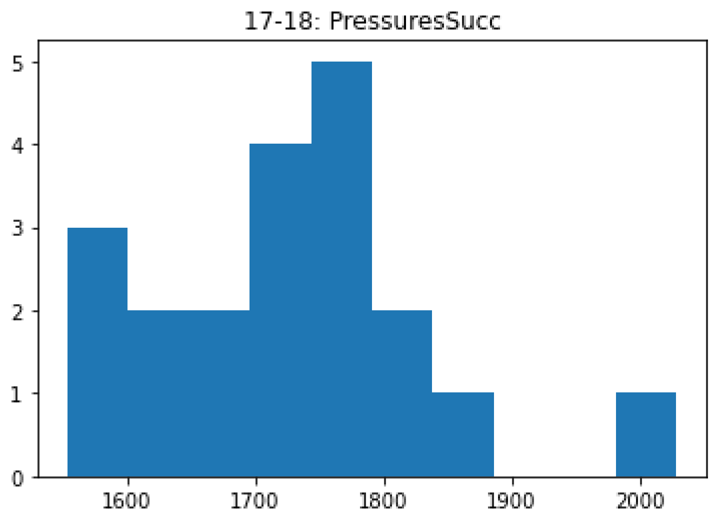
Mean 5826.9



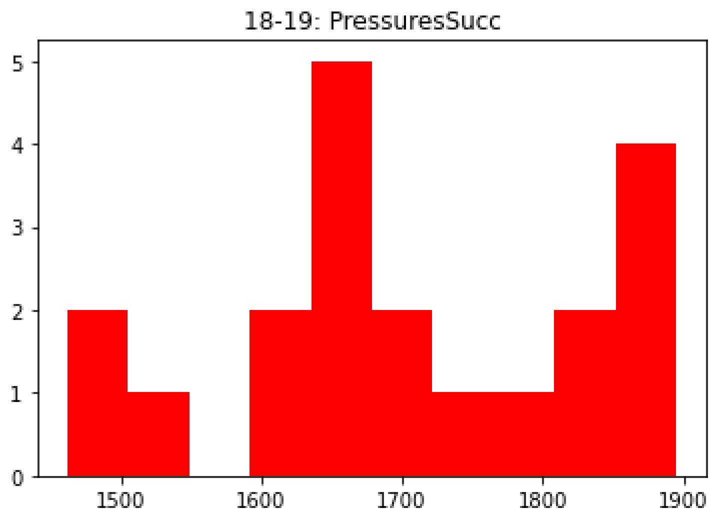
Mean 5931.25



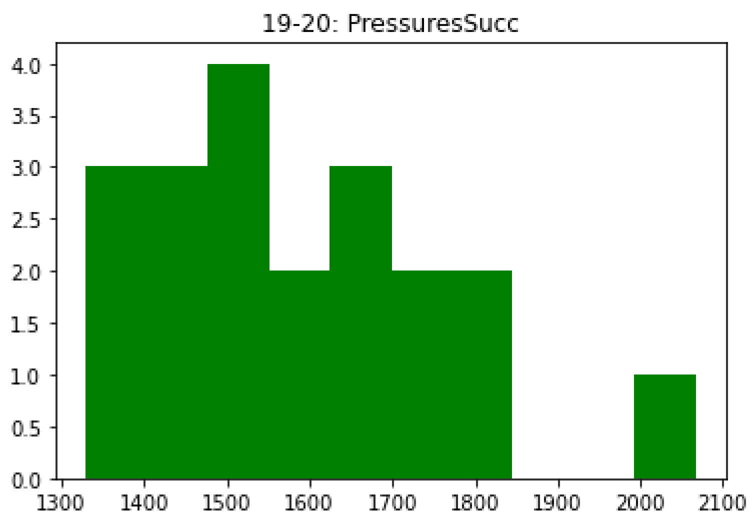
Mean 5435.65
ANOVA Values
F_onewayResult(statistic=0.3258742406865952, pvalue=0.7232319670414316)
PressuresSucc



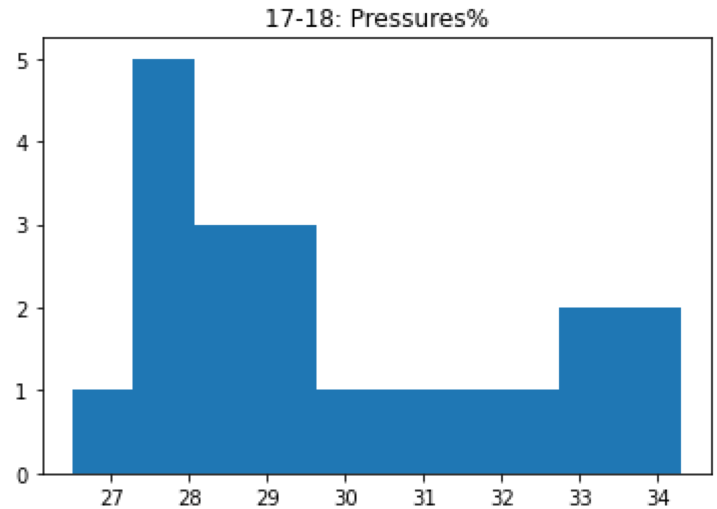
Mean 1726.95



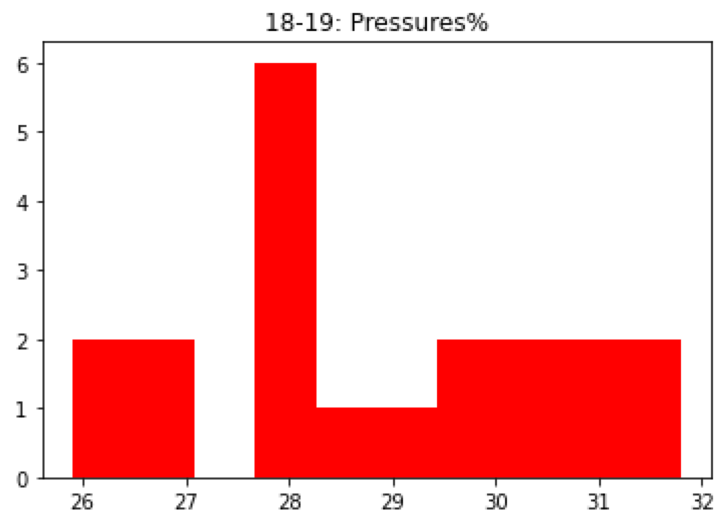

```
Mean 1705.35
```



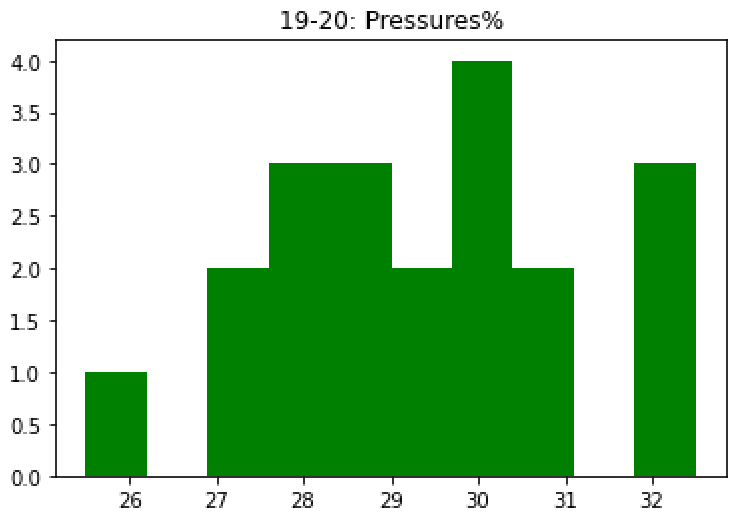

```
Mean 1589.3
ANOVA Values
F_onewayResult(statistic=0.1955945745620415, pvalue=0.8228951796960114)
Pressures%
```



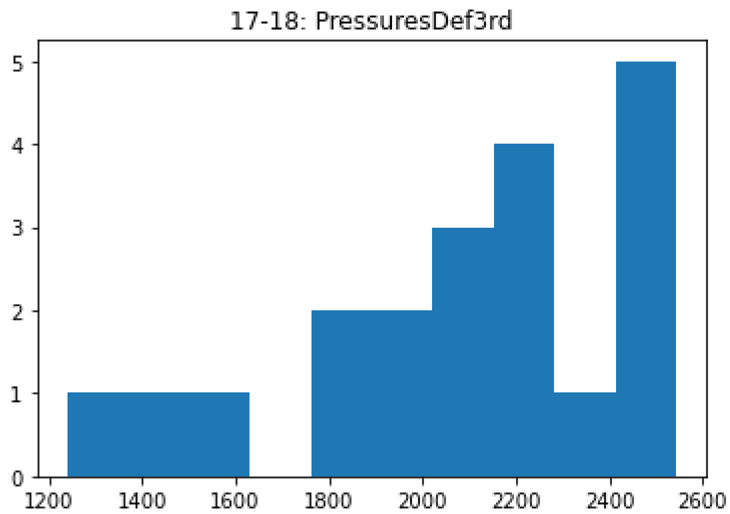
Mean 29.765000000000008



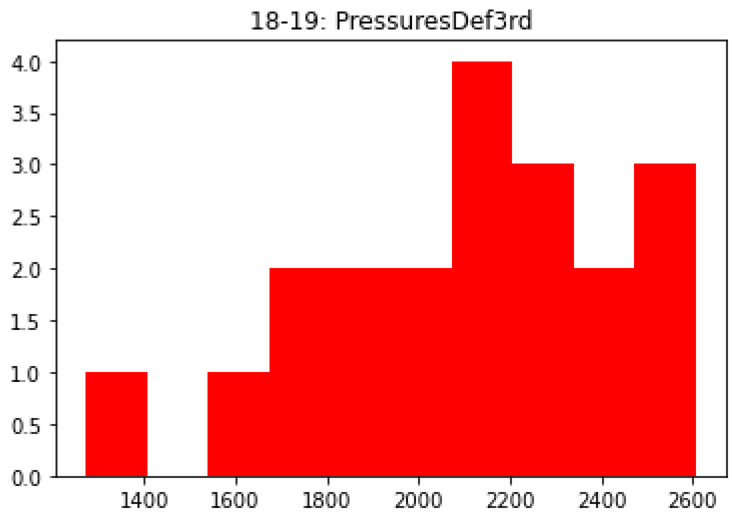
Mean 28.800000000000004



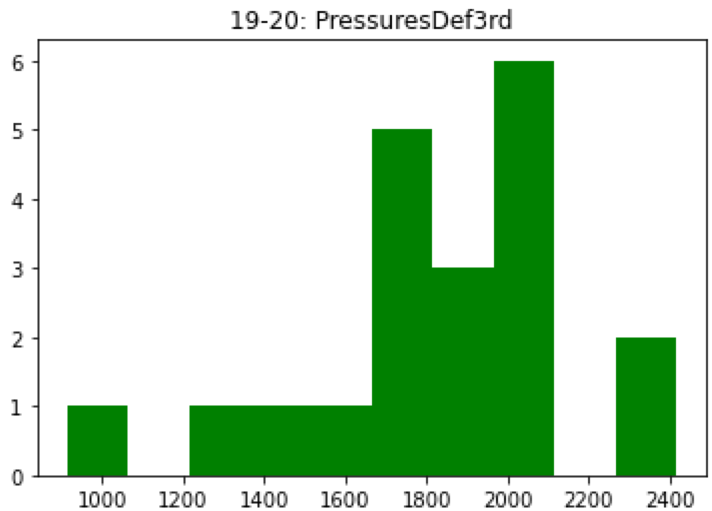
Mean 29.28
ANOVA Values
F_onewayResult(statistic=1.5729767453231718, pvalue=0.2162970439582328)
PressuresDef3rd



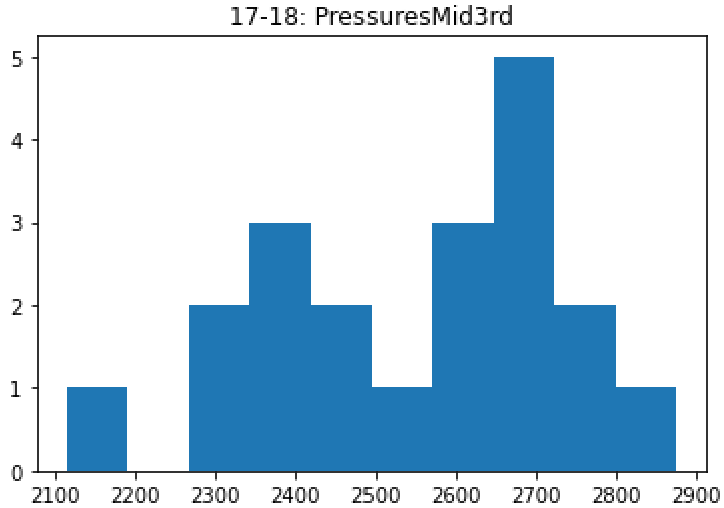
Mean 2076.85



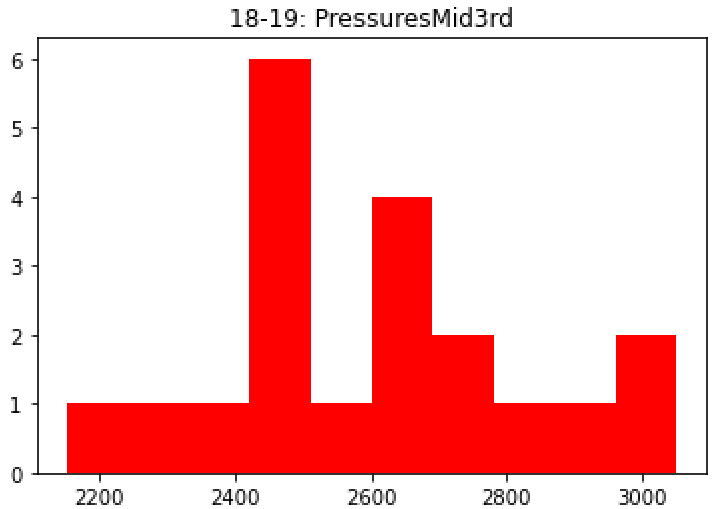

```
Mean 2098.6
```



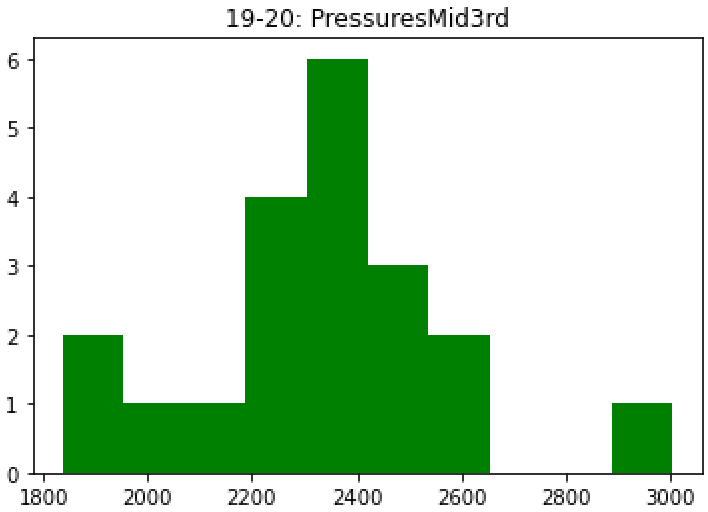

```
Mean 1828.15
ANOVA Values
F_onewayResult(statistic=0.025664293332984412, pvalue=0.9746734911923046)
PressuresMid3rd
```



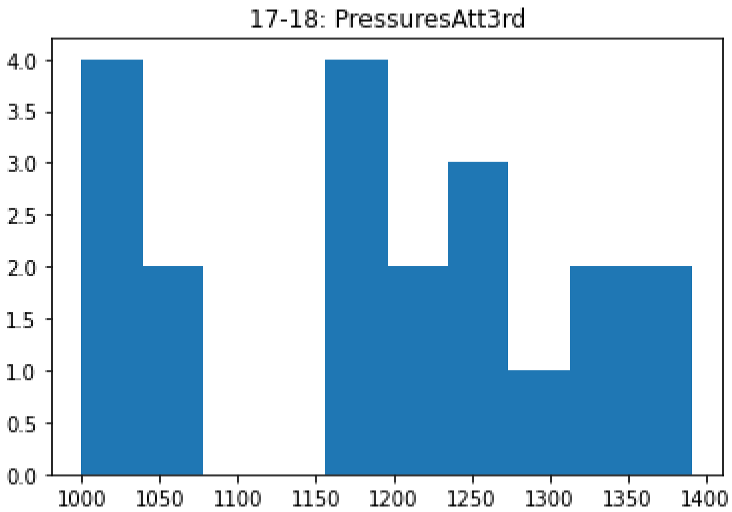
Mean 2554.3



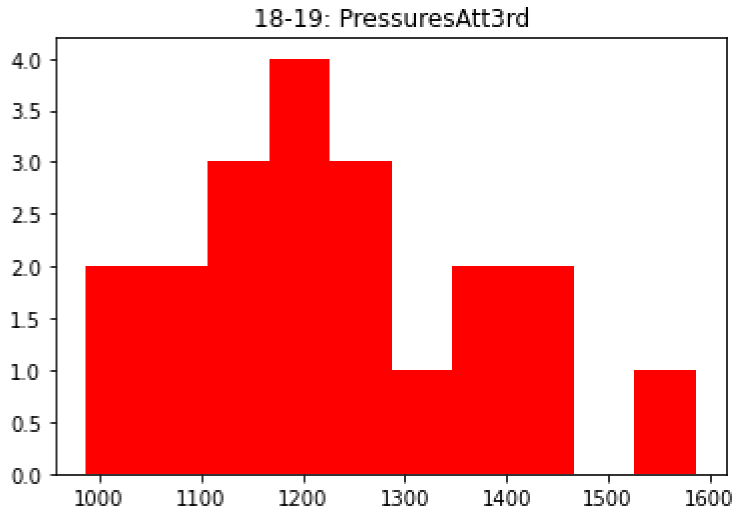
Mean 2601.8



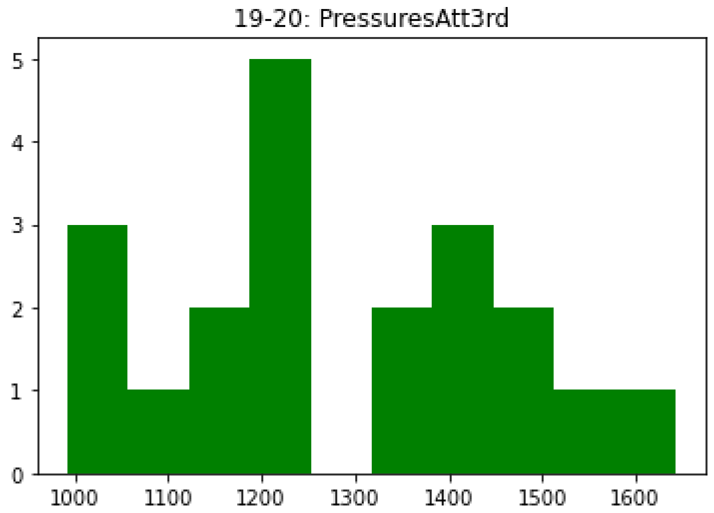

```
Mean 2322.1
ANOVA Values
F_onewayResult(statistic=0.3260399261867283, pvalue=0.7231135026929335)
PressuresAtt3rd
```



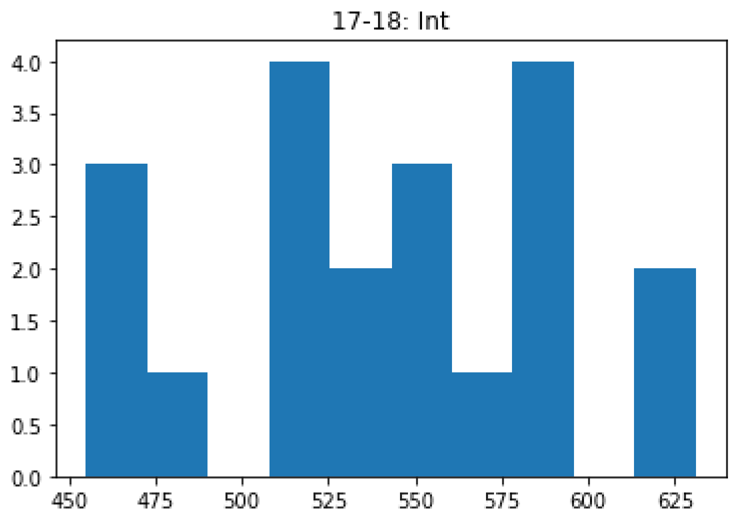

```
Mean 1195.75
```



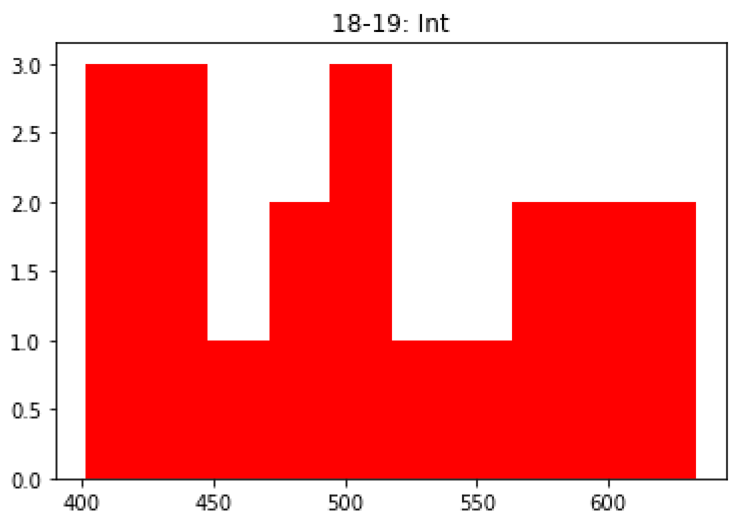

```
Mean 1230.85
```



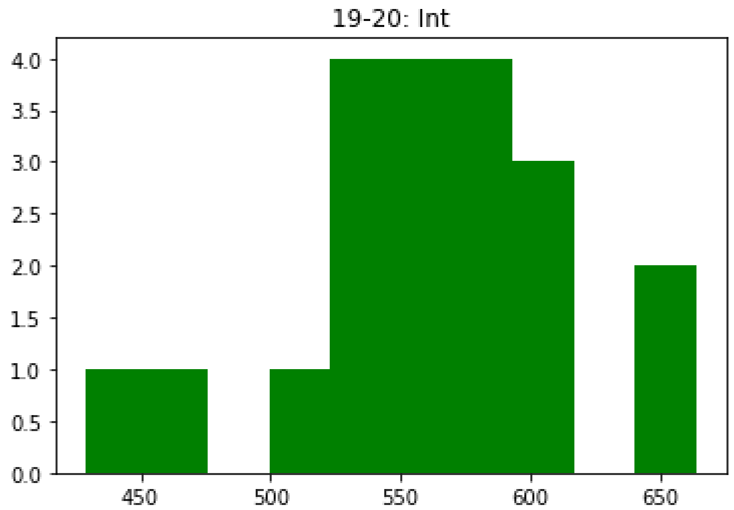

```
Mean 1285.4
ANOVA Values
F_onewayResult(statistic=0.39992709863813114, pvalue=0.672235145269355)
Int
```



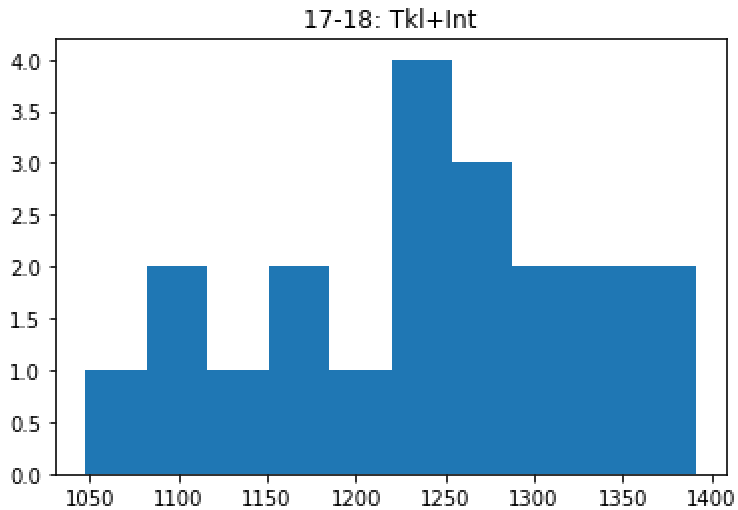
Mean 543.75



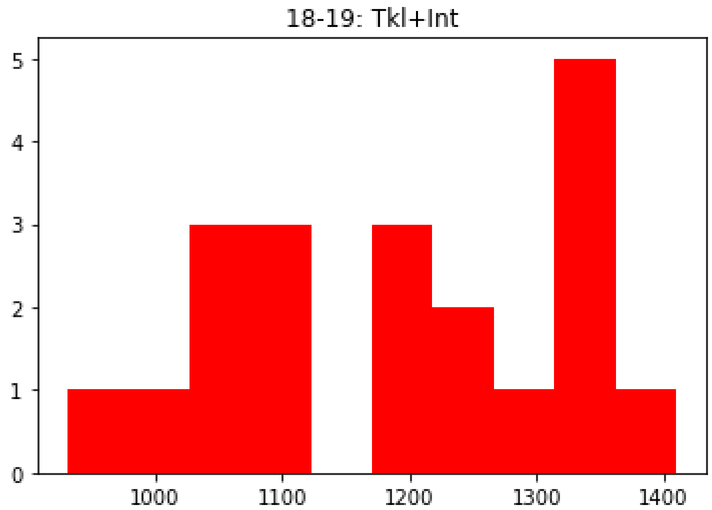
Mean 509.45



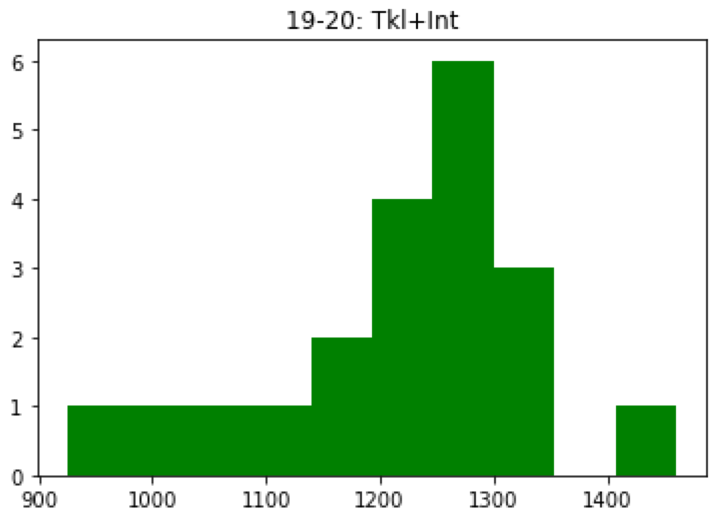

```
Mean 561.9
ANOVA Values
F_onewayResult(statistic=1.7414161771525367, pvalue=0.1844666880156552)
Tkl+Int
```



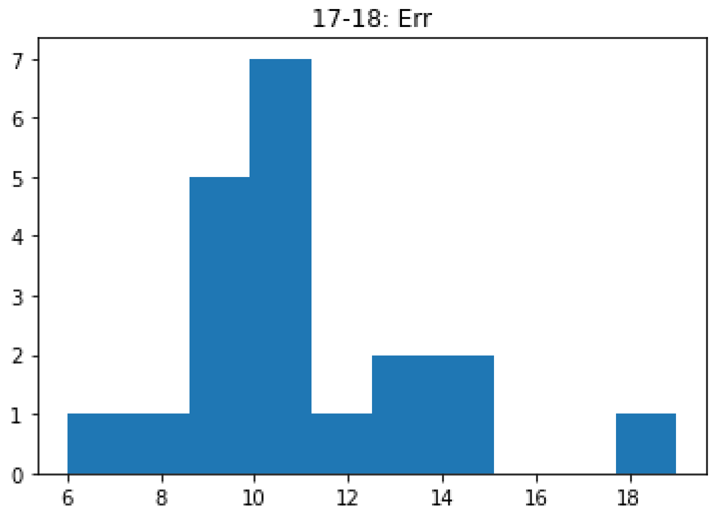

```
Mean 1240.4
```



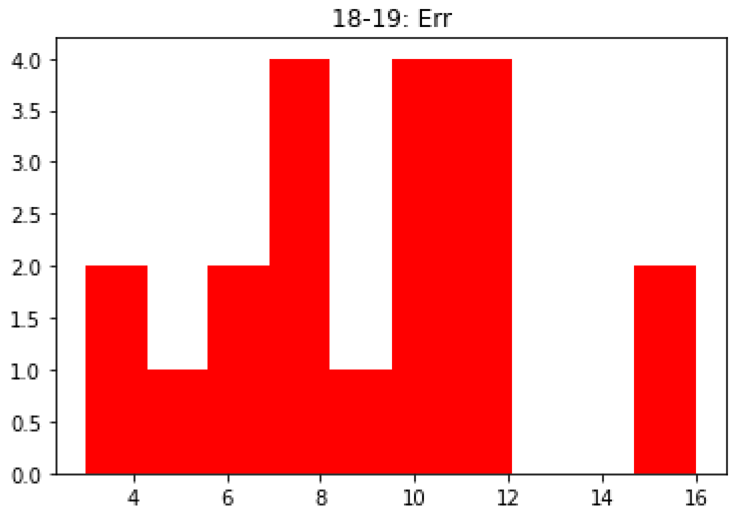
Mean 1192.75



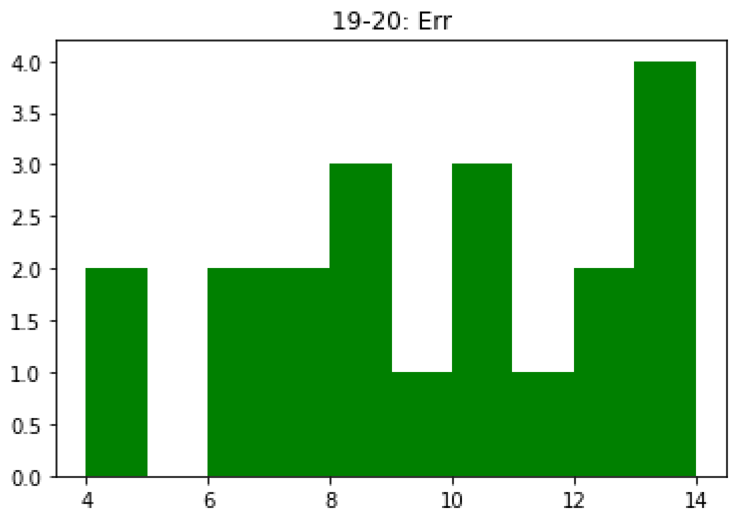

```
Mean 1219.3
ANOVA Values
F_onewayResult(statistic=0.9981377434714589, pvalue=0.3749146675708669)
Err
```



Mean 10.85



Mean 8.9




```
Mean 9.3
ANOVA Values
F_onewayResult(statistic=2.4401756311745344, pvalue=0.09620096604255621)
```

In [ ]:

```
for x in range(len(statnames)):
    print(statnames[x])
    print("17-18 Mean")
    print(np.mean(data1718.loc[:,statnames[x]]))
    print("18-19 Mean")
    print(np.mean(data1819.loc[:,statnames[x]]))
    print("19-20 Mean")
    print(np.mean(data1920.loc[:,statnames[x]]))
    print("Probability")
    print(f_oneway(data1718.loc[:,statnames[x]], data1819.loc[:,statnames[x]],data1819.loc[:,statnames[x]]).pvalue)
```

```
#Pl
17-18 Mean
26.45
18-19 Mean
25.4
19-20 Mean
27.1
Probability
0.3807404291169192
TacklesTkl
17-18 Mean
696.65
```

18-19 Mean
683.3
19-20 Mean
657.4
Probability
0.7926945755671473
TacklesTklW
17-18 Mean
422.55
18-19 Mean
413.95
19-20 Mean
393.45
Probability
0.6888521227568817
TacklesDef3rd
17-18 Mean
351.95
18-19 Mean
349.6
19-20 Mean
319.0
Probability
0.9877576172012333
TacklesMid3rd
17-18 Mean
264.8
18-19 Mean
253.3
19-20 Mean
251.35
Probability
0.3339536030136585
TacklesAtt3rd
17-18 Mean
79.9
18-19 Mean
80.4
19-20 Mean
87.05
Probability
0.9933261300560847
VsDribblesTkl
17-18 Mean
228.05

18-19 Mean
213.4
19-20 Mean
235.3
Probability
0.05593503742377694
VsDribblesAtt
17-18 Mean
634.75
18-19 Mean
594.15
19-20 Mean
575.95
Probability
0.054593104048497404
VsDribblesTkl%
17-18 Mean
35.955
18-19 Mean
35.98
19-20 Mean
40.81
Probability
0.9990775854882242
VsDribblesPast
17-18 Mean
406.7
18-19 Mean
380.75
19-20 Mean
340.65
Probability
0.111495965036905
PressuresPress
17-18 Mean
5826.9
18-19 Mean
5931.25
19-20 Mean
5435.65
Probability
0.7232319670414316
PressuresSucc
17-18 Mean
1726.95

18-19 Mean
1705.35
19-20 Mean
1589.3
Probability
0.8228951796960114
Pressures%
17-18 Mean
29.765000000000008
18-19 Mean
28.800000000000004
19-20 Mean
29.28
Probability
0.2162970439582328
PressuresDef3rd
17-18 Mean
2076.85
18-19 Mean
2098.6
19-20 Mean
1828.15
Probability
0.9746734911923046
PressuresMid3rd
17-18 Mean
2554.3
18-19 Mean
2601.8
19-20 Mean
2322.1
Probability
0.7231135026929335
PressuresAtt3rd
17-18 Mean
1195.75
18-19 Mean
1230.85
19-20 Mean
1285.4
Probability
0.672235145269355
Int
17-18 Mean
543.75

```
18-19 Mean
509.45
19-20 Mean
561.9
Probability
0.1844666880156552
Tkl+Int
17-18 Mean
1240.4
18-19 Mean
1192.75
19-20 Mean
1219.3
Probability
0.3749146675708669
Err
17-18 Mean
10.85
18-19 Mean
8.9
19-20 Mean
9.3
Probability
0.09620096604255621
```

In [ ]:

```
import seaborn as sns
```

In [ ]:

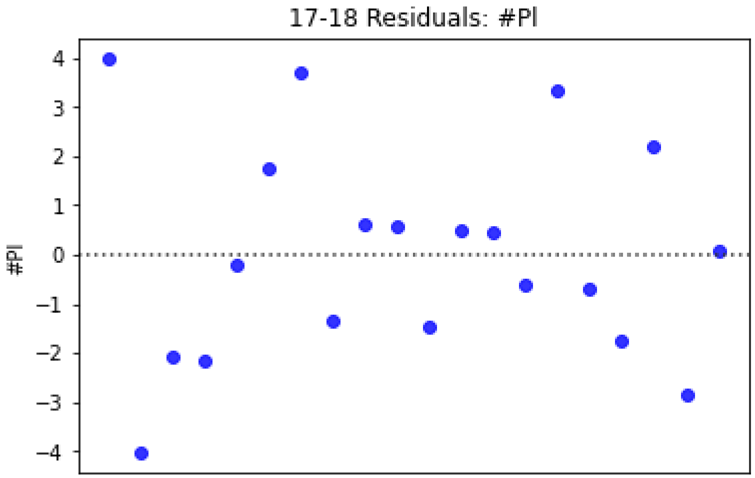
```
def listofones(x):
    l = []
    for b in range(x):
        l.append(b)
    return l
```

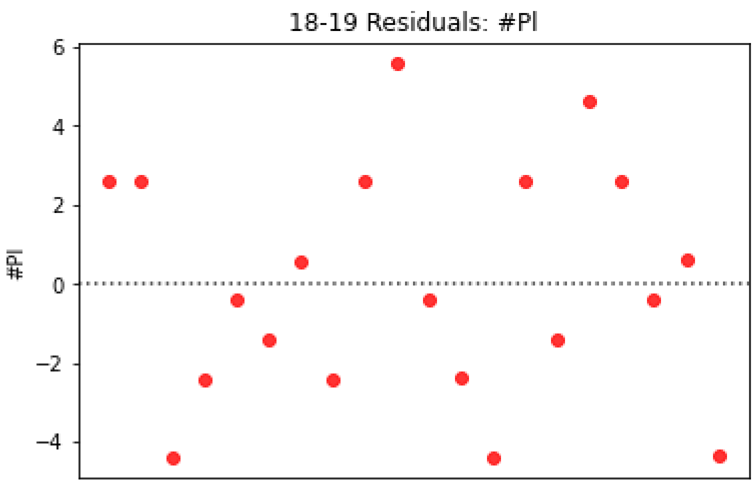
In [ ]:

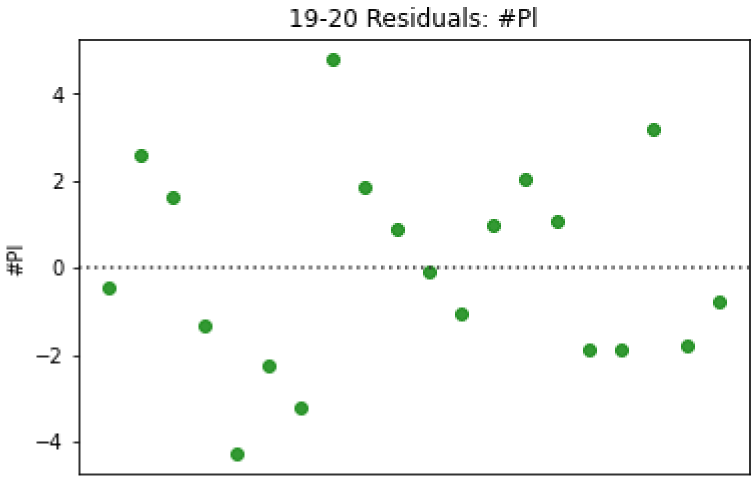
```
for x in range(len(statnames)):
    print(statnames[x])
    sns.residplot(x = listofones(len(data1718.loc[:,statnames[x]])),y = data1718.loc[:,statnames[x]],color = "blue")
    plt.title("17-18 Residuals: " + statnames[x])
    ax = plt.gca()
    ax.axes.xaxis.set_ticks([])
    plt.show()

    sns.residplot(x = listofones(len(data1819.loc[:,statnames[x]])),y = data1819.loc[:,statnames[x]], color = "red")
```

```
    plt.title("18-19 Residuals: " + statnames[x])
    ax = plt.gca()
    ax.axes.xaxis.set_ticks([])
    plt.show()

    sns.residplot(x = listofones(len(data1920.loc[:,statnames[x]])),y = data1920.loc[:,statnames[x]], color = "green")
    plt.title("19-20 Residuals: " + statnames[x])
    ax = plt.gca()
    ax.axes.xaxis.set_ticks([])
    plt.show()
```

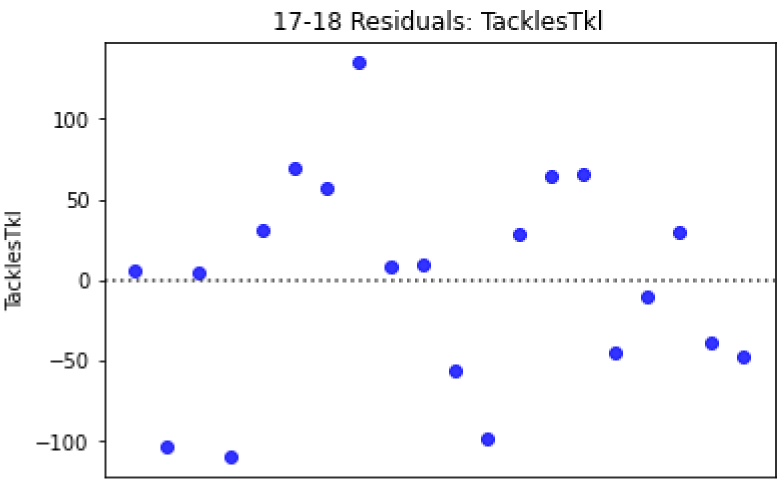
#Pl

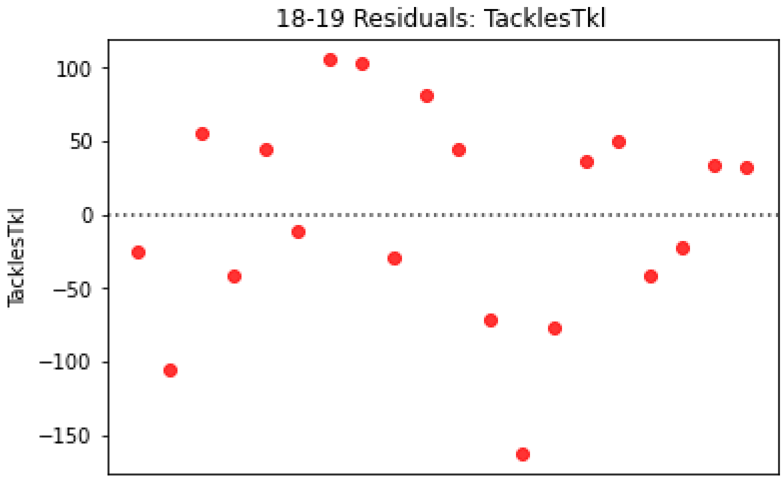


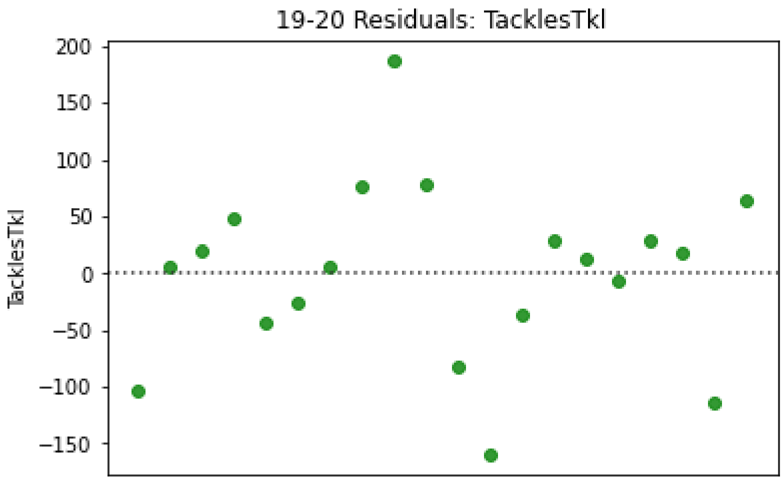



19-20 Residuals: #Pl
#Pl
4
2
0
−2
−4

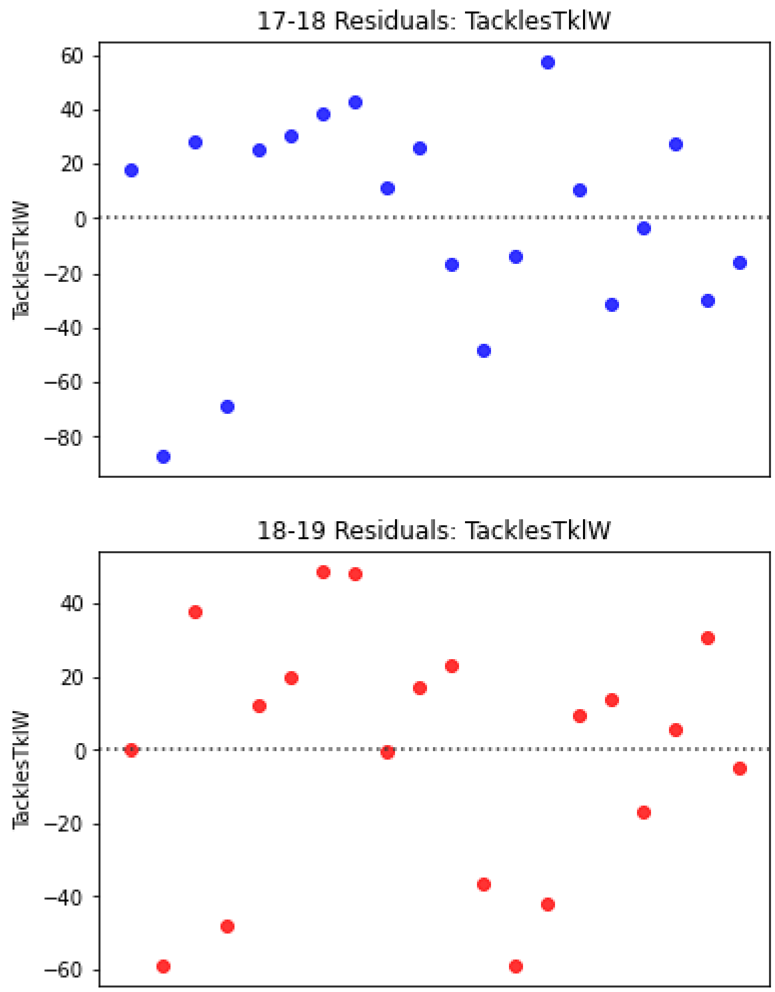
TacklesTkl

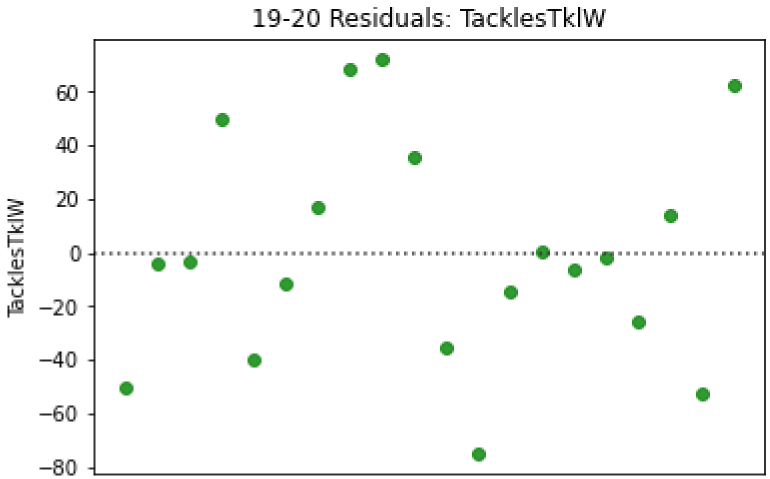

17-18 Residuals: TacklesTkl
TacklesTkl
100
50
0
−50
−100



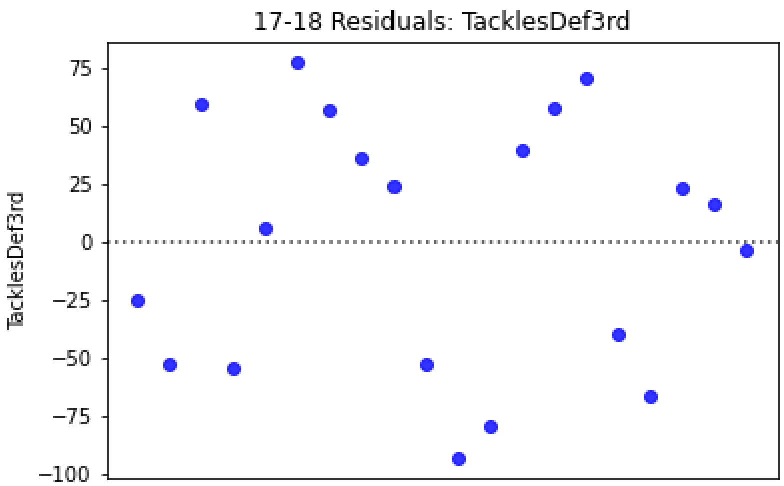


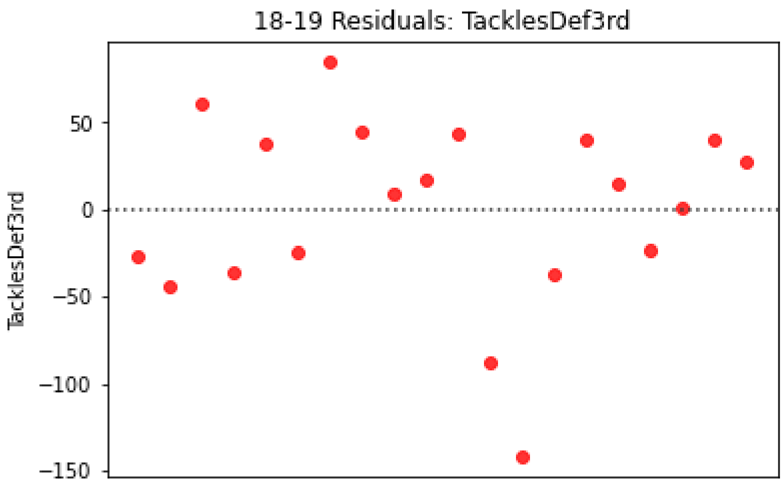
TacklesTklW

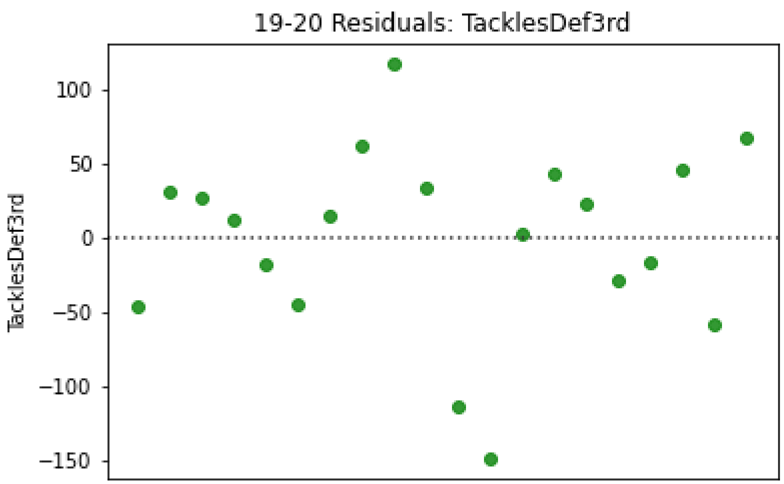

17-18 Residuals: TacklesTklW
TacklesTklW
60
40
20
0
−20
−40
−60
−80
18-19 Residuals: TacklesTklW
TacklesTklW
40
20
0
−20
−40
−60



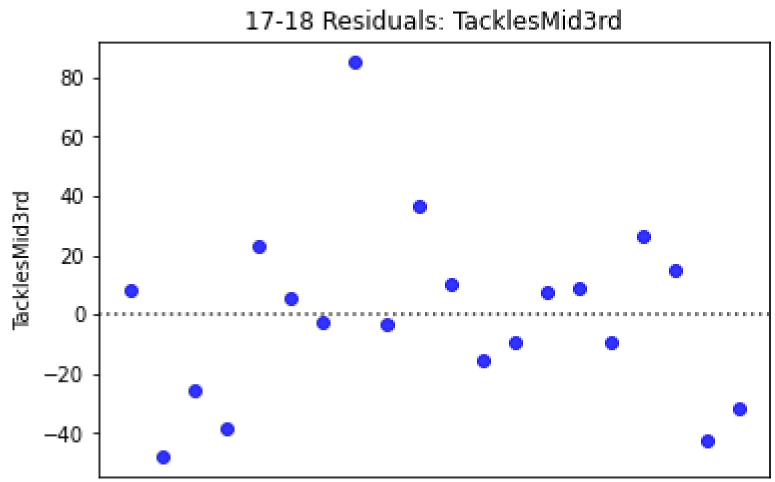
`TacklesDef3rd`

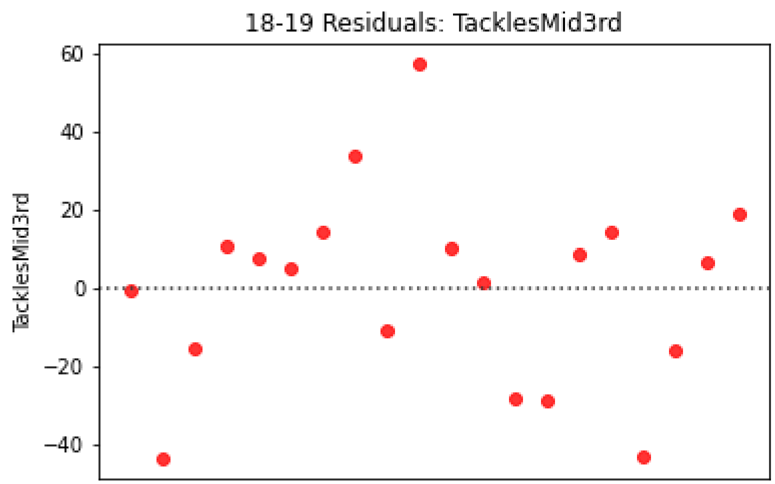


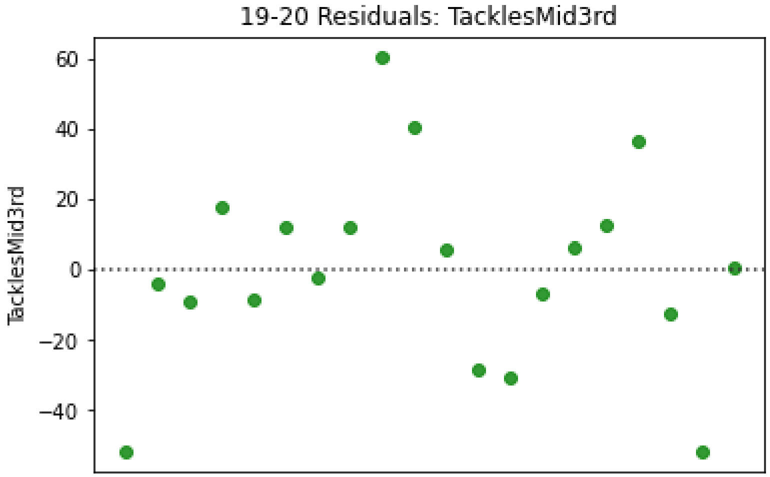




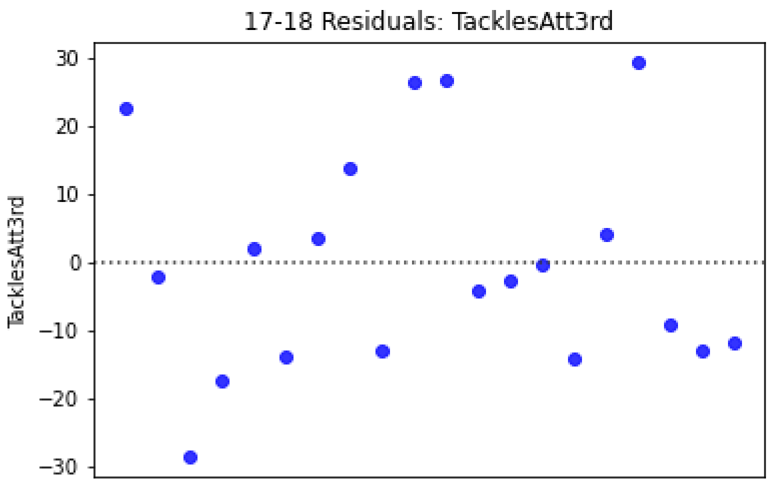
TacklesMid3rd

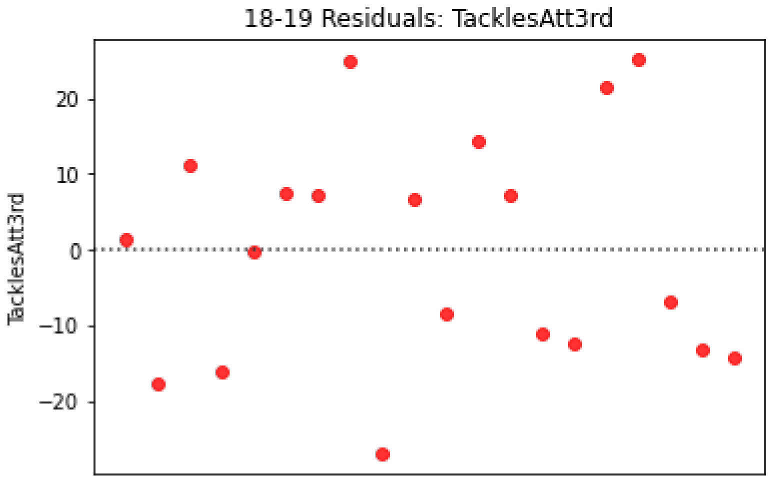

17-18 Residuals: TacklesMid3rd
TacklesMid3rd
80
60
40
20
0
−20
−40

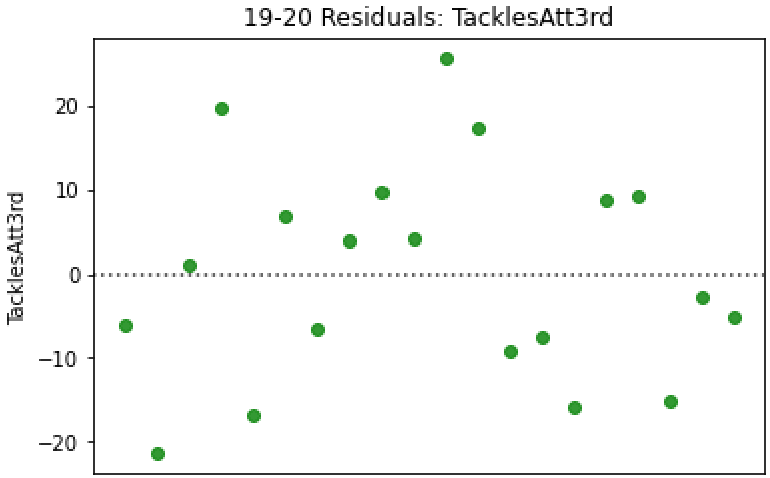

18-19 Residuals: TacklesMid3rd
TacklesMid3rd
60
40
20
0
−20
−40



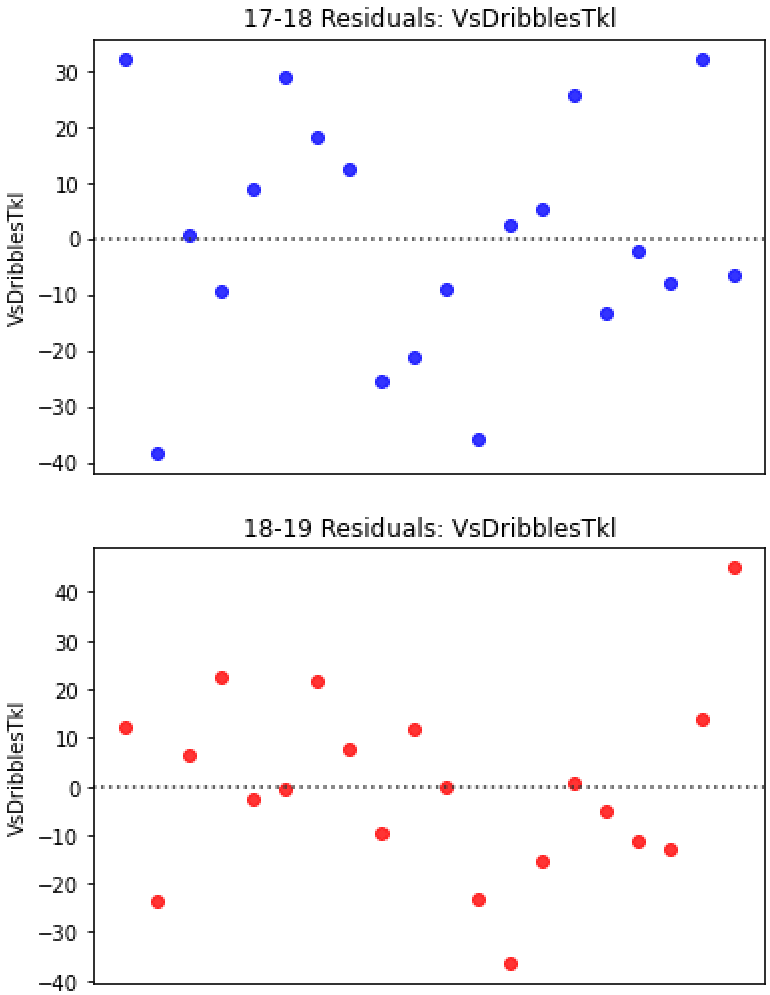
TacklesAtt3rd

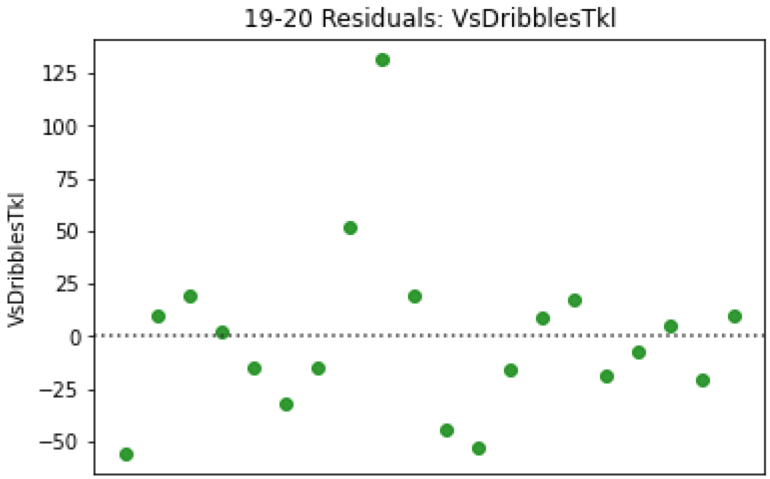




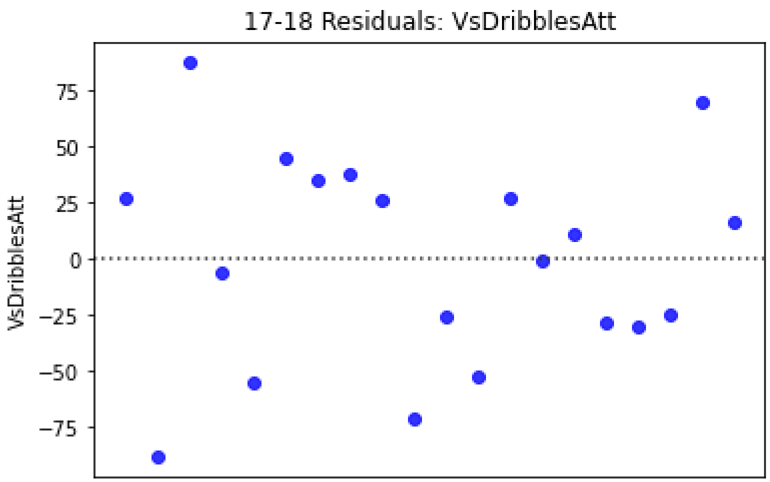


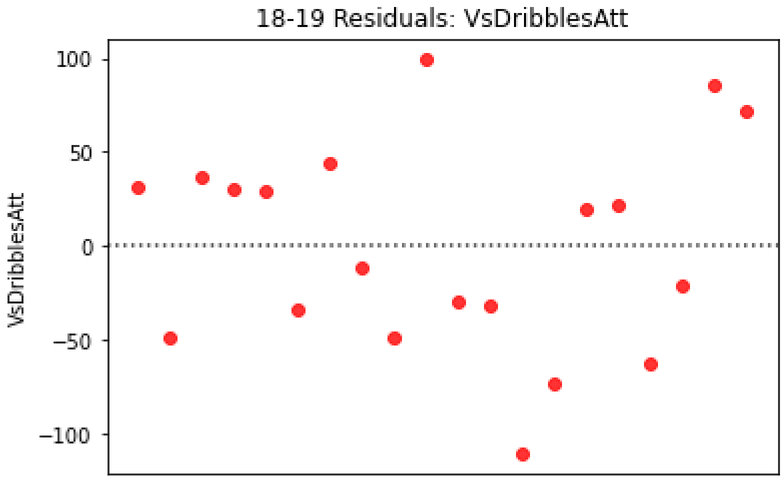
VsDribblesTkl

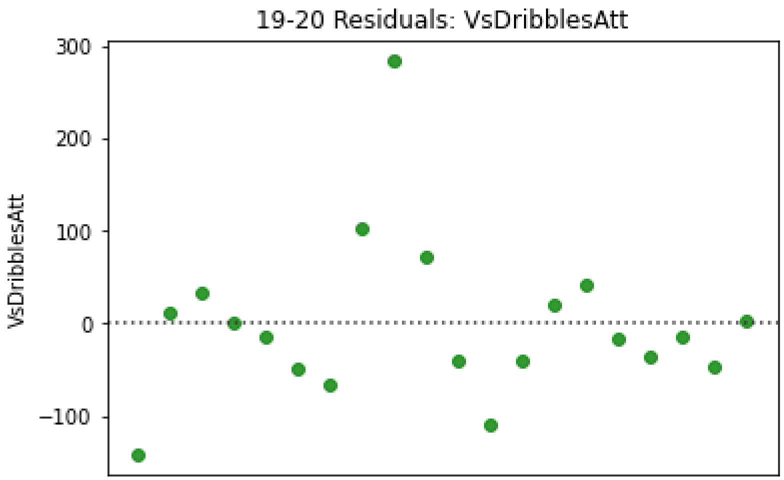

17-18 Residuals: VsDribblesTkl
VsDribblesTkl
30
20
10
0
−10
−20
−30
−40
18-19 Residuals: VsDribblesTkl
VsDribblesTkl
40
30
20
10
0
−10
−20
−30
−40



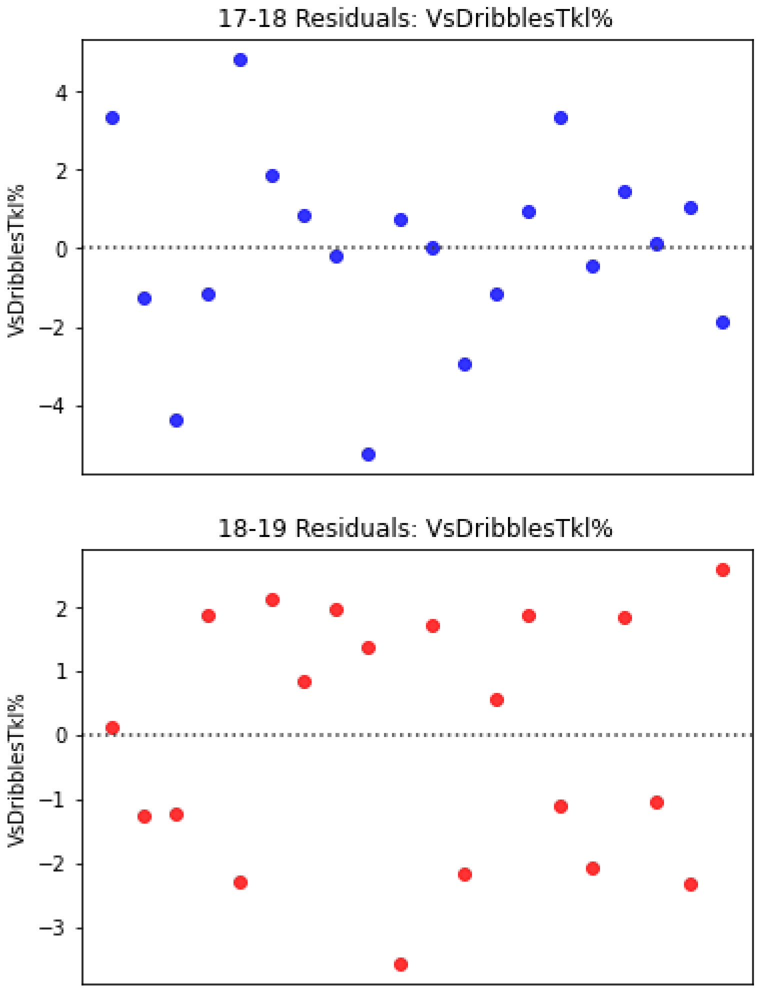
VsDribblesAtt



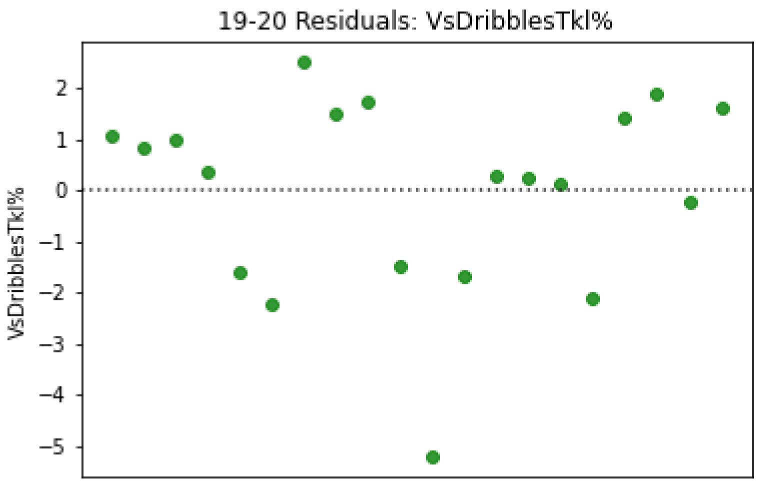




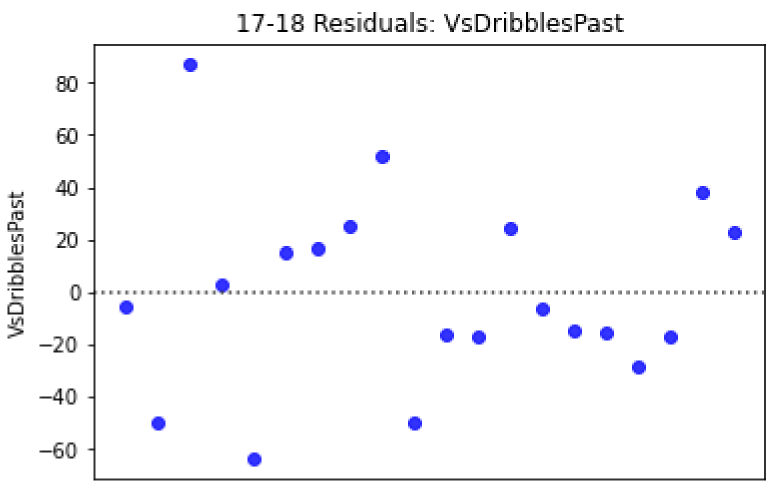
VsDribblesTkl%

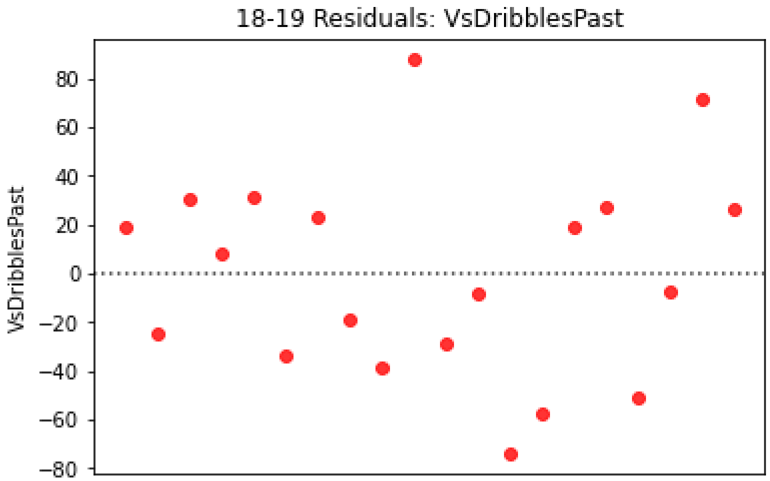


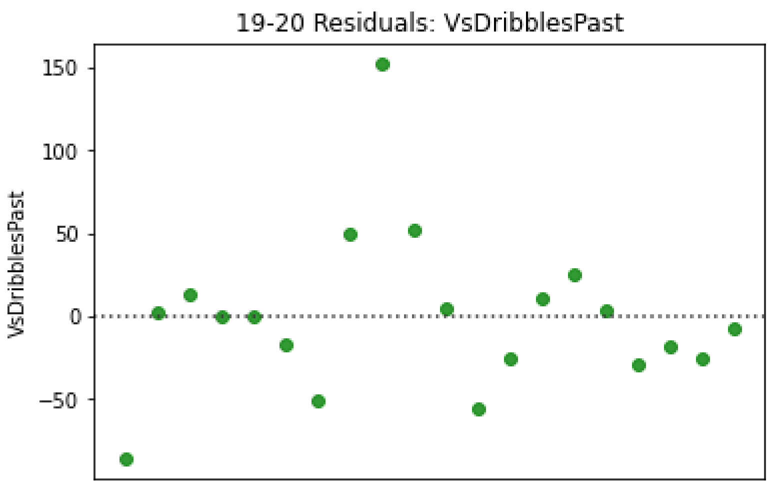
17-18 Residuals: VsDribblesTkl%
VsDribblesTkl%
4
2
0
−2
−4
18-19 Residuals: VsDribblesTkl%
VsDribblesTkl%
2
1
0
−1
−2
−3



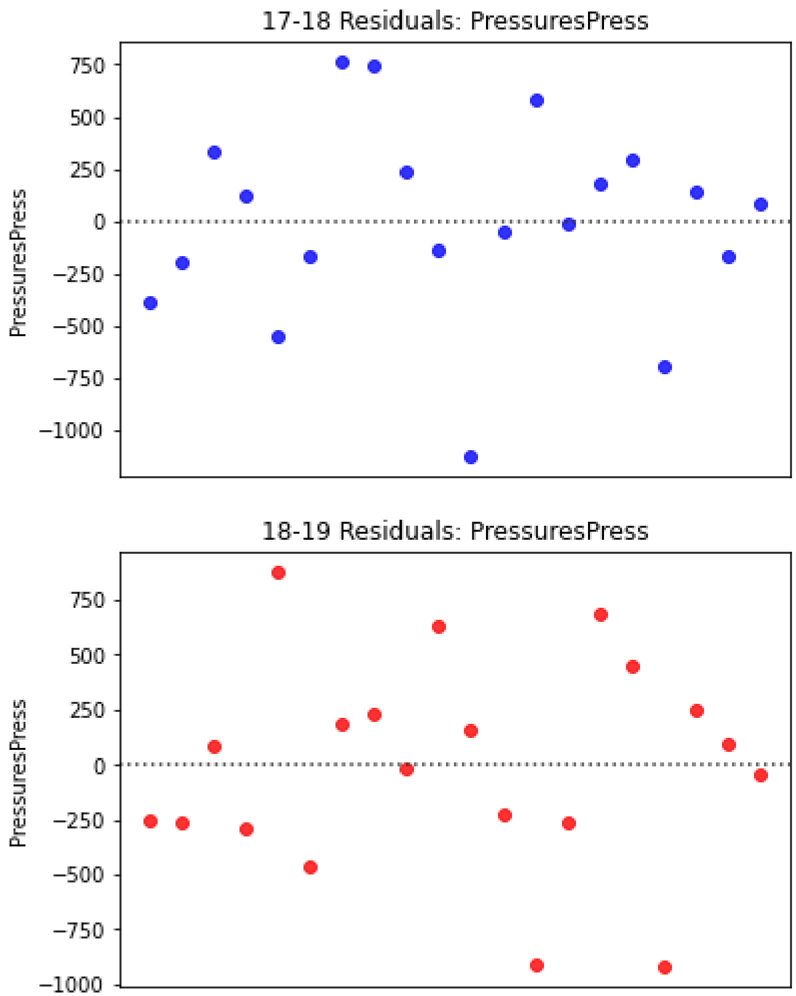
VsDribblesPast

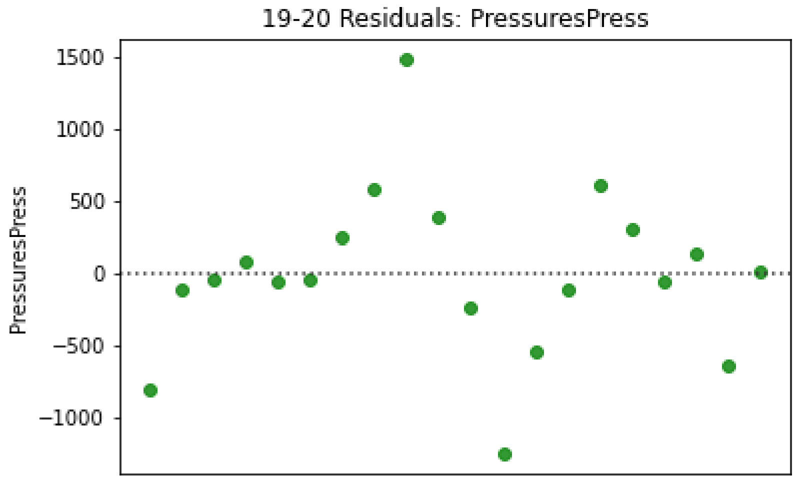




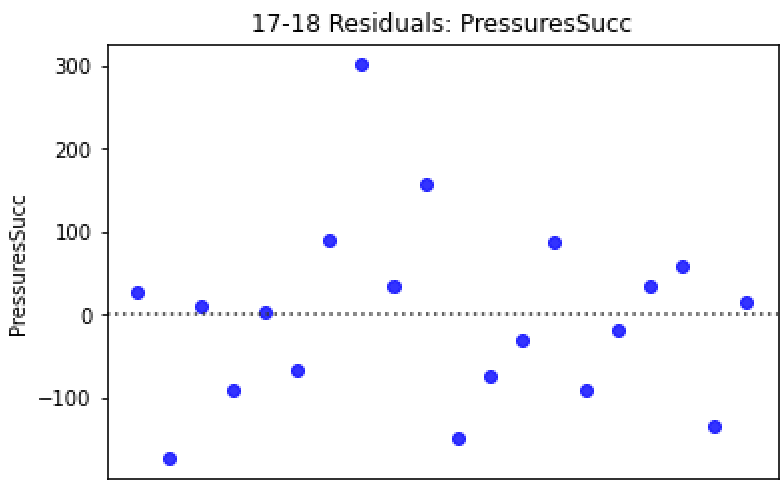


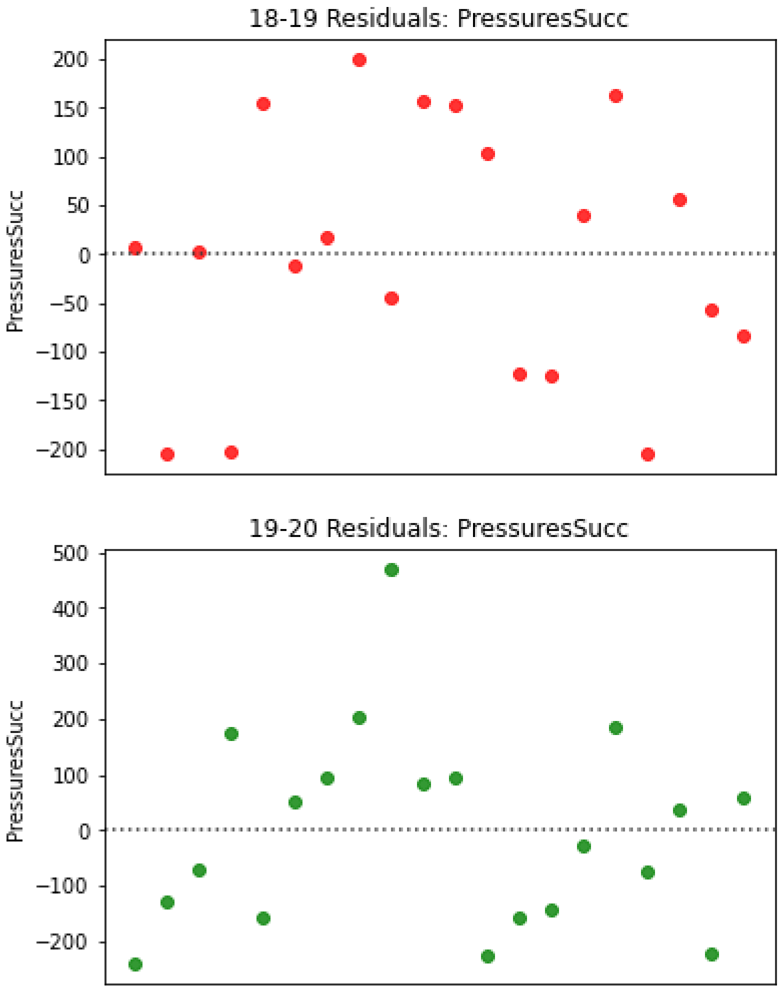
PressuresPress


17-18 Residuals: PressuresPress
PressuresPress
750
500
250
0
−250
−500
−750
−1000
18-19 Residuals: PressuresPress
PressuresPress
750
500
250
0
−250
−500
−750
−1000

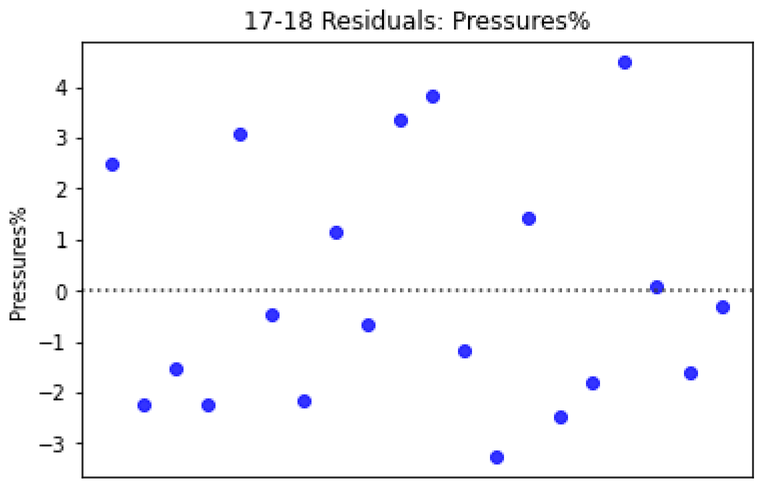


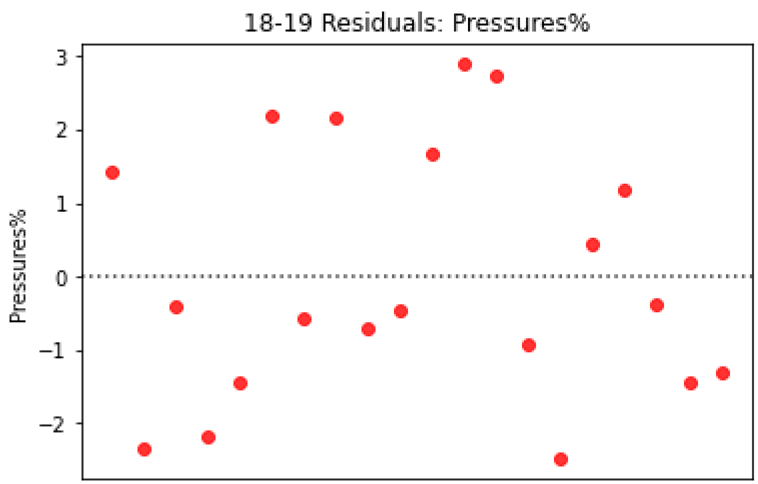
PressuresSucc

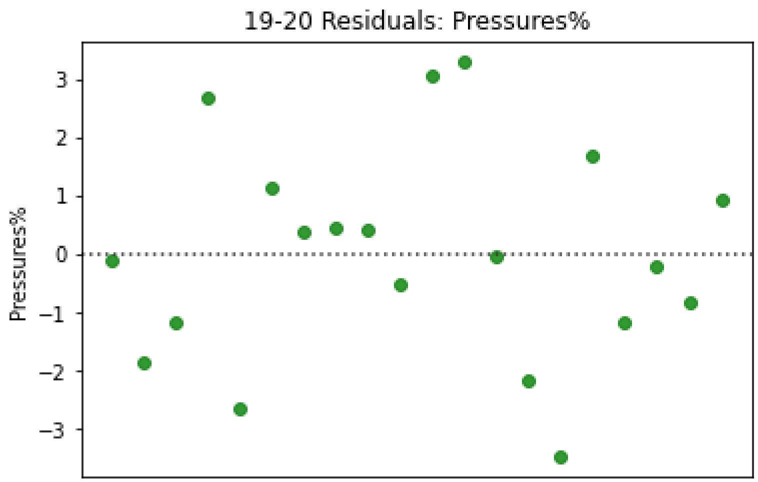



18-19 Residuals: PressuresSucc
PressuresSucc
200
150
100
50
0
−50
−100
−150
−200
19-20 Residuals: PressuresSucc
PressuresSucc
500
400
300
200
100
0
−100
−200

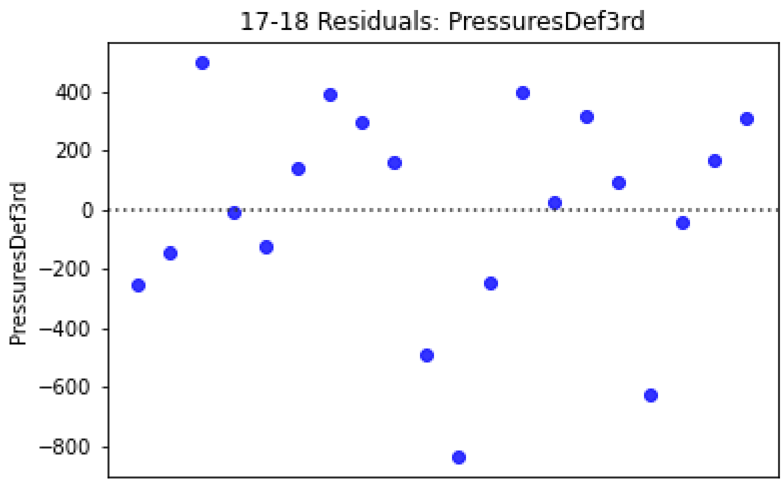
Pressures%

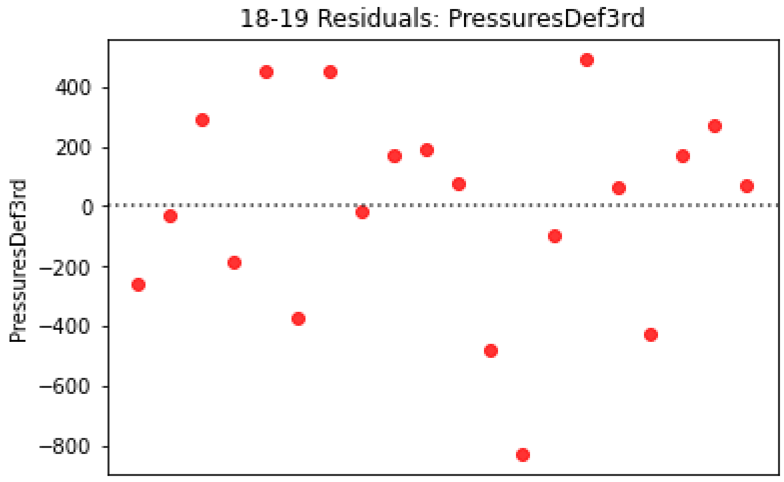

17-18 Residuals: Pressures%
Pressures%
4
3
2
1
0
−1
−2
−3

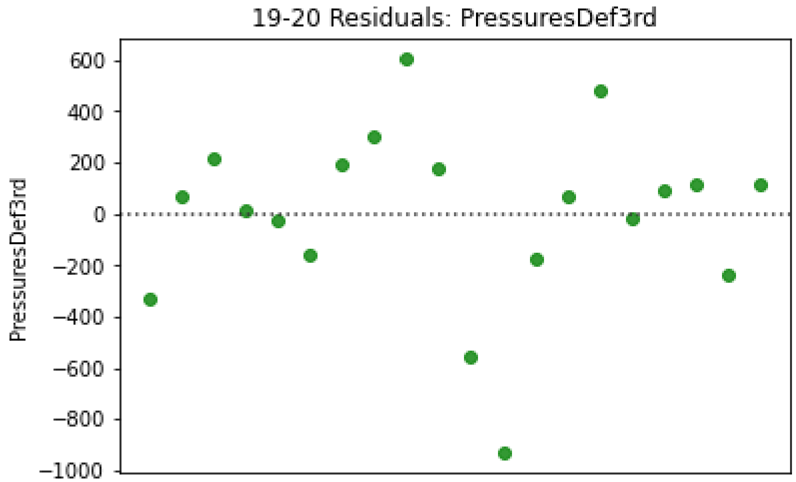

18-19 Residuals: Pressures%
Pressures%
3
2
1
0
−1
−2



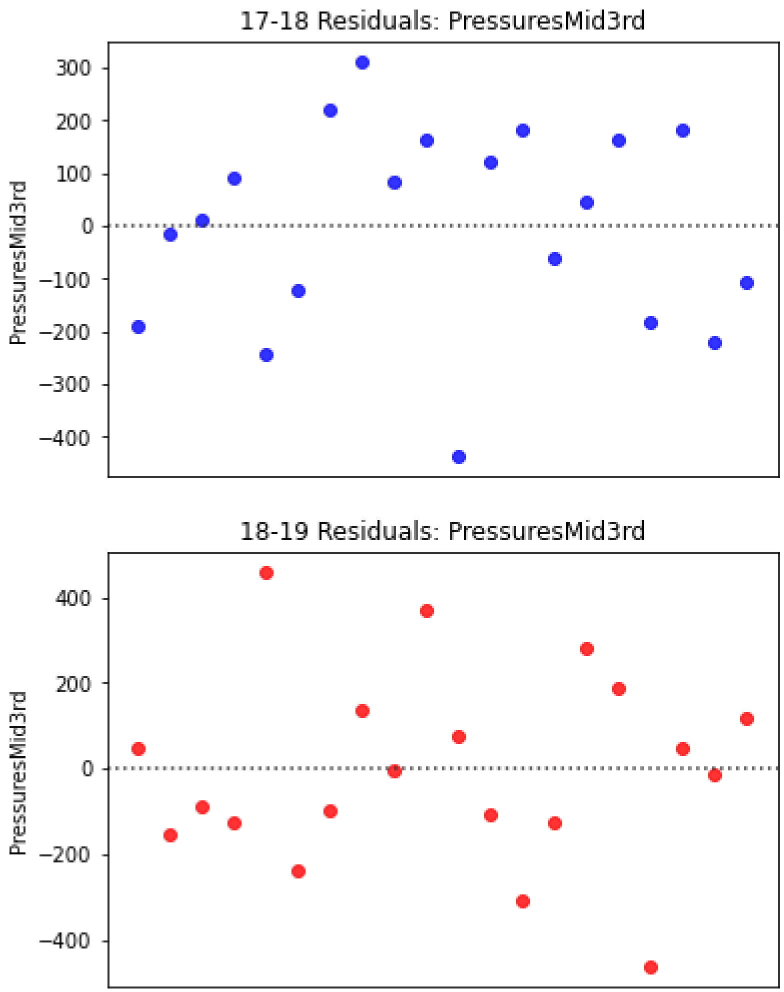
PressuresDef3rd

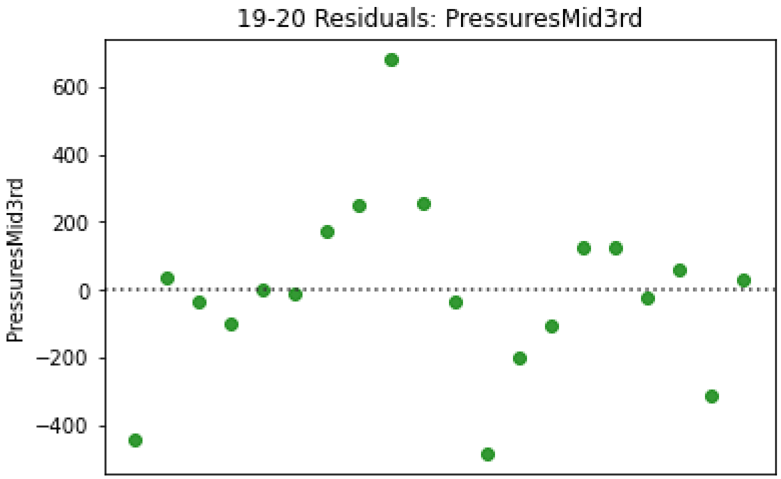




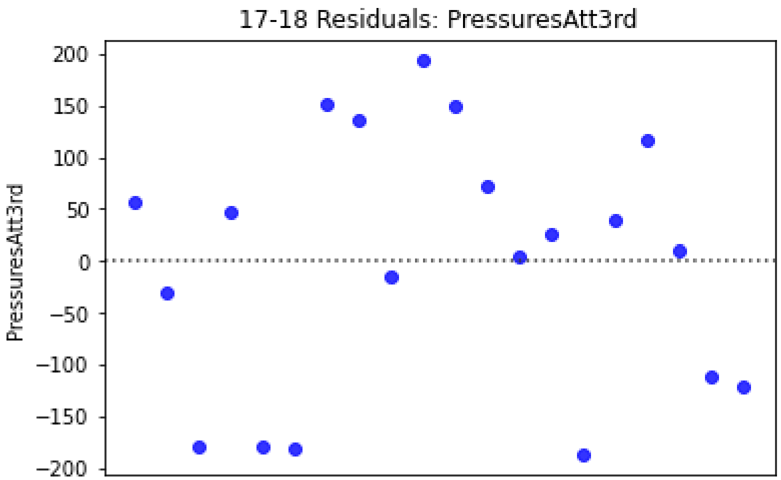


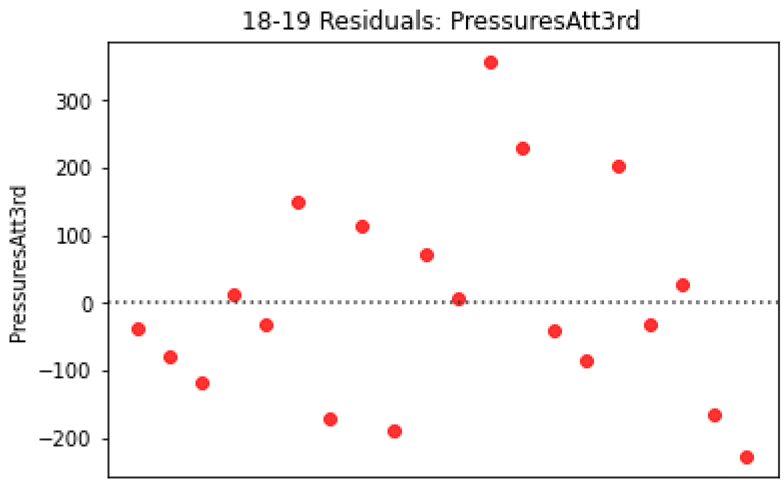
PressuresMid3rd

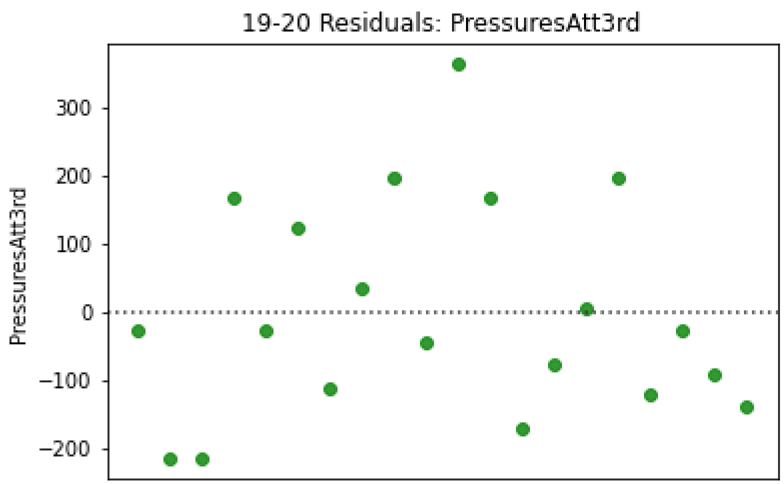

17-18 Residuals: PressuresMid3rd
PressuresMid3rd
300
200
100
0
−100
−200
−300
−400
18-19 Residuals: PressuresMid3rd
PressuresMid3rd
400
200
0
−200
−400



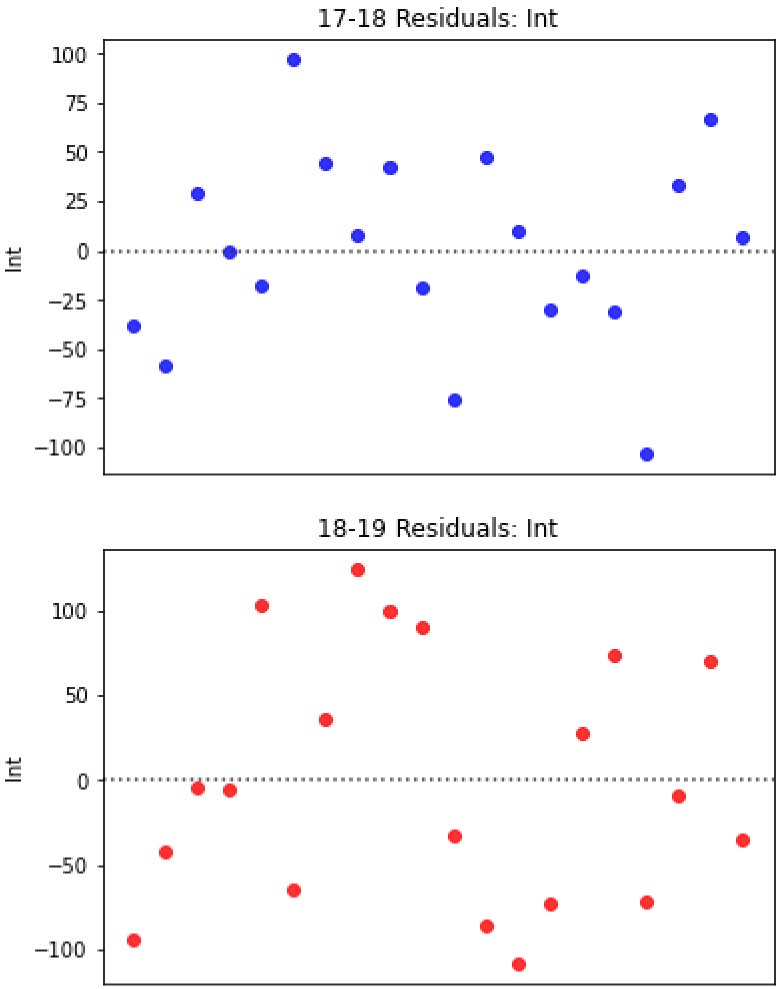
PressuresAtt3rd

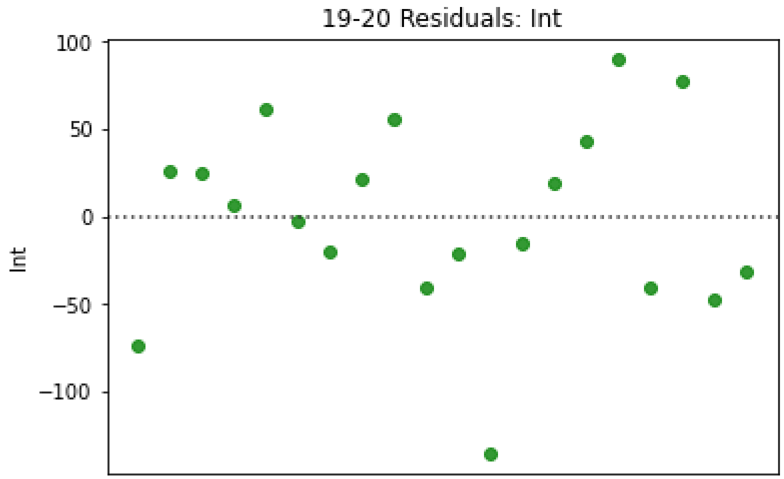



18-19 Residuals: PressuresAtt3rd
PressuresAtt3rd
300
200
100
0
−100
−200

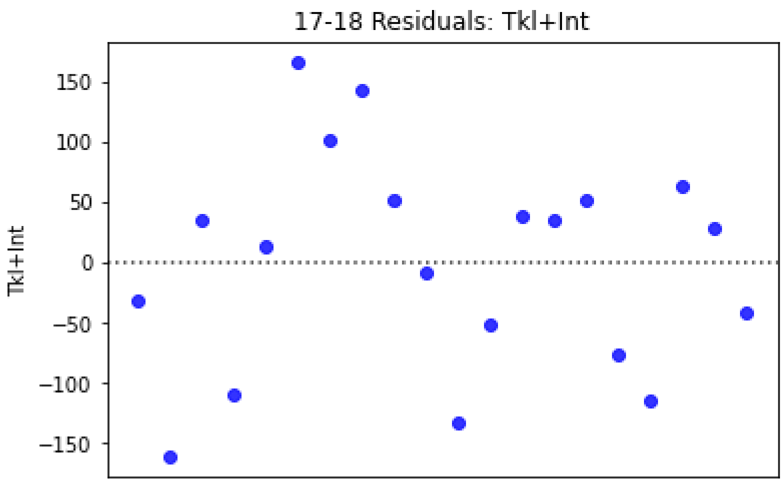

19-20 Residuals: PressuresAtt3rd
PressuresAtt3rd
300
200
100
0
−100
−200

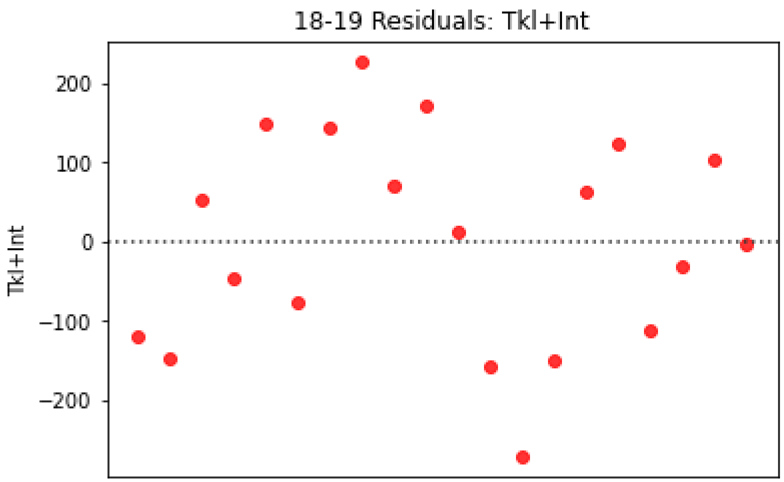
Int

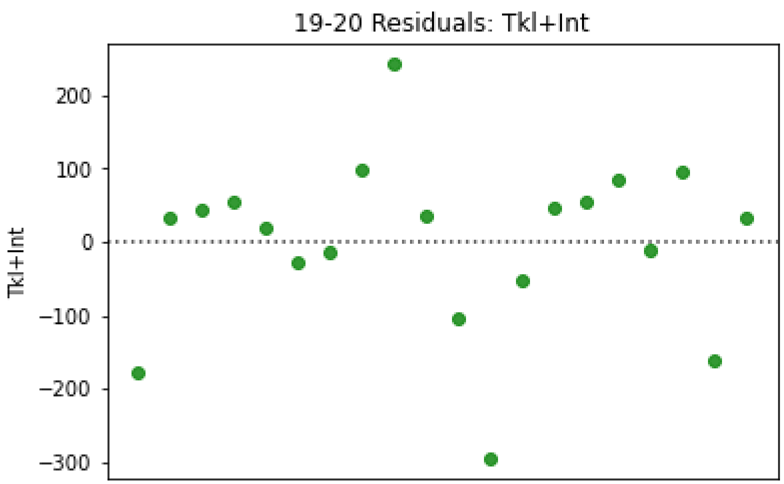

17-18 Residuals: Int
Int
100
75
50
25
0
−25
−50
−75
−100
18-19 Residuals: Int
Int
100
50
0
−50
−100


19-20 Residuals: Int
100
50
0
−50
−100
Int

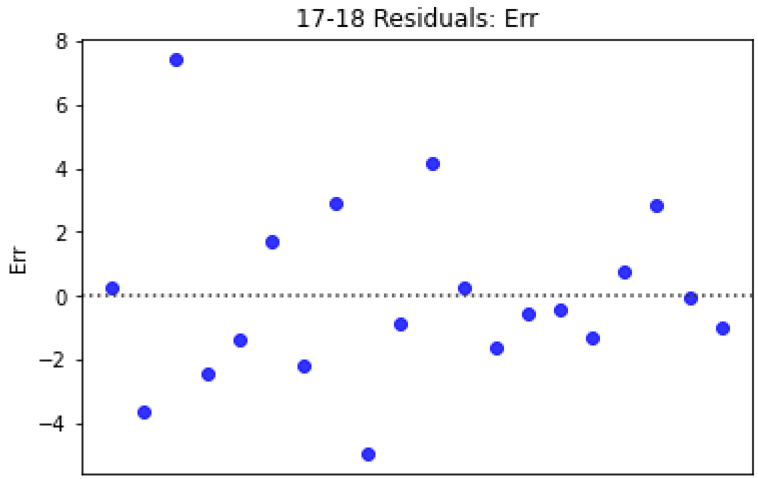
Tkl+Int


17-18 Residuals: Tkl+Int
150
100
50
0
−50
−100
−150
Tkl+Int

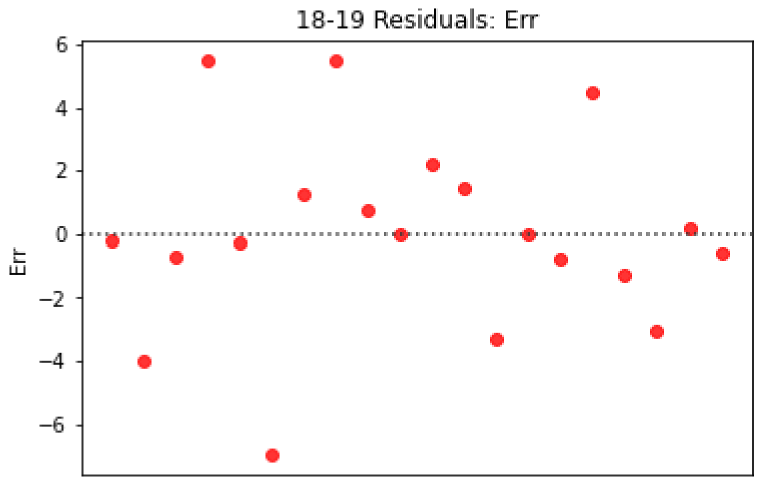

18-19 Residuals: Tkl+Int
Tkl+Int
200
100
0
−100
−200


19-20 Residuals: Tkl+Int
Tkl+Int
200
100
0
−100
−200
−300

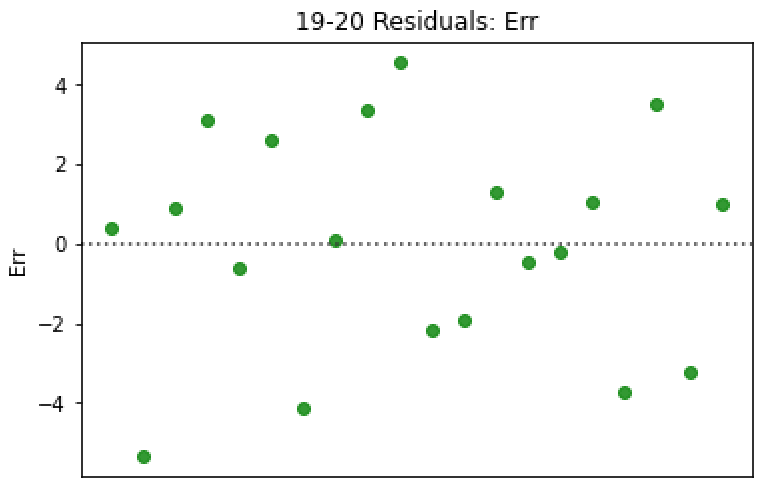
Err



17-18 Residuals: Err
Err
8
6
4
2
0
−2
−4



18-19 Residuals: Err
Err
6
4
2
0
−2
−4
−6



In [ ]: