In [42]:

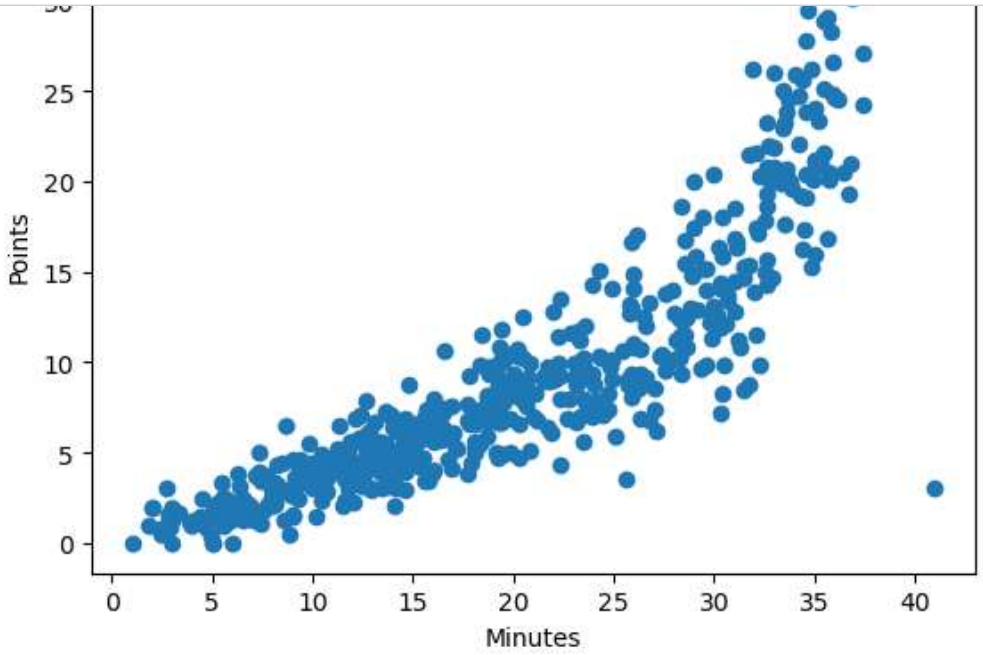
```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

# Load the customer data into a DataFrame
nba_df = pd.read_csv('C:/Users/srken/Downloads/nba_players_22-23/nba_pl

# Remove duplicate players- i.e. players who changed teams during the se
nba_df.drop_duplicates(subset = "Player", keep = "first", inplace = Tru

# Check the first 10 rows
pd.set_option('display.max_columns', None)
nba_df.head(10)
```

Out[42]:

Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	...	FT%	ORB	DRB
1	Precious Achiuwa	C	23	TOR	55	12	20.7	3.6	7.3	0.485	...	0.702	1.8	4.1
2	Steven Adams	C	29	MEM	42	42	27.0	3.7	6.3	0.597	...	0.364	5.1	6.5
3	Bam Adebayo	C	25	MIA	75	75	34.6	8.0	14.9	0.540	...	0.806	2.5	6.7
4	Ochai Agbaji	SG	22	UTA	59	22	20.5	2.8	6.5	0.427	...	0.812	0.7	1.3
5	Santi Aldama	PF	22	MEM	77	20	21.8	3.2	6.8	0.470	...	0.750	1.1	3.7
6	Nickeil Alexander-Walker	SG	24	TOT	59	3	15.0	2.2	5.0	0.444	...	0.667	0.3	1.5
7	Grayson Allen	SG	27	MIL	72	70	27.4	3.4	7.7	0.440	...	0.905	0.8	2.4
8	Jarrett Allen	C	24	CLE	68	68	32.6	5.9	9.2	0.644	...	0.733	3.3	6.5
9	Jose Alvarado	PG	24	NOP	61	10	21.5	3.3	8.0	0.411	...	0.813	0.5	1.9
10	Kyle Anderson	PF	29	MIN	69	46	28.4	3.7	7.2	0.509	...	0.735	1.0	4.4

10 rows × 29 columns

In [43]:
```
plt.scatter(nba_df["MP"],
            nba_df["PTS"])

plt.xlabel("Minutes")
plt.ylabel("Points")
```



In [60]:
```
nba_df2 = nba_df[nba_df['G'] >= 41]
nba_df2.head(5)
```

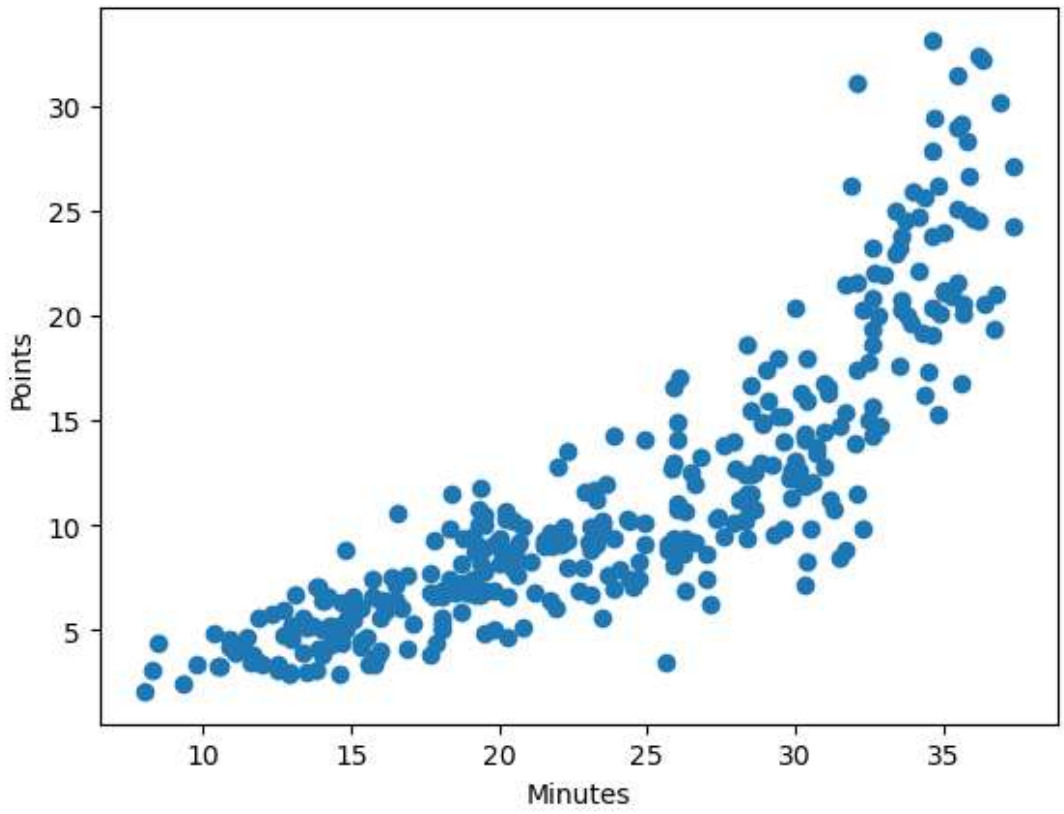
Out[60]:

Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	...	FT%	ORB	DRB	T
1	Precious Achiuwa	C	23	TOR	55	12	20.7	3.6	7.3	0.485	...	0.702	1.8	4.1	
2	Steven Adams	C	29	MEM	42	42	27.0	3.7	6.3	0.597	...	0.364	5.1	6.5	1
3	Bam Adebayo	C	25	MIA	75	75	34.6	8.0	14.9	0.540	...	0.806	2.5	6.7	
4	Ochai Agbaji	SG	22	UTA	59	22	20.5	2.8	6.5	0.427	...	0.812	0.7	1.3	
5	Santi Aldama	PF	22	MEM	77	20	21.8	3.2	6.8	0.470	...	0.750	1.1	3.7	

5 rows × 29 columns

In [61]:
```
plt.scatter(nba_df2["MP"],
            nba_df2["PTS"])

plt.xlabel("Minutes")
plt.ylabel("Points")
```

Out[61]: Text(0, 0.5, 'Points')

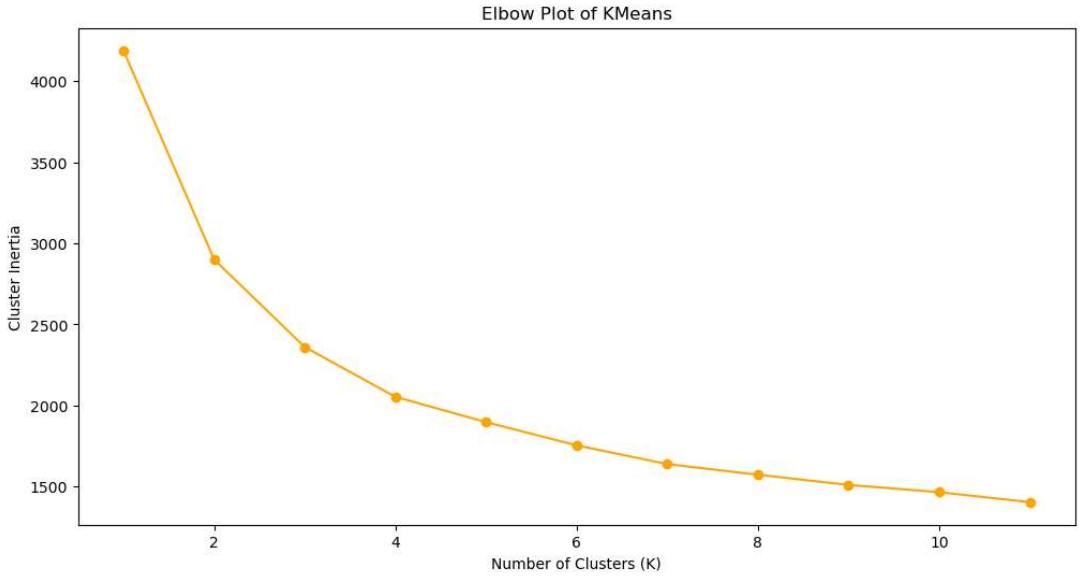


In [79]:
```
relevant_cols = ["PTS", "3PA", "3P%", "2PA", "2P%",
                 "FTA", "FT%", "TRB", "AST", "STL", "BLK", "TOV"]

nba_df3 = nba_df2[relevant_cols]
```

In [80]:
```
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

scaler.fit(nba_df3)

scaled_data = scaler.transform(nba_df3)
```

In [81]:

```
def find_best_clusters(df, maximum_K):

    clusters_centers = []
    k_values = []

    for k in range(1, maximum_K):

        kmeans_model = KMeans(n_clusters = k)
        kmeans_model.fit(df)

        clusters_centers.append(kmeans_model.inertia_)
        k_values.append(k)


    return clusters_centers, k_values

def generate_elbow_plot(clusters_centers, k_values):

    figure = plt.subplots(figsize = (12, 6))
    plt.plot(k_values, clusters_centers, 'o-', color = 'orange')
    plt.xlabel("Number of Clusters (K)")
    plt.ylabel("Cluster Inertia")
    plt.title("Elbow Plot of KMeans")
    plt.show()

clusters_centers, k_values = find_best_clusters(scaled_data, 12)

generate_elbow_plot(clusters_centers, k_values)
```

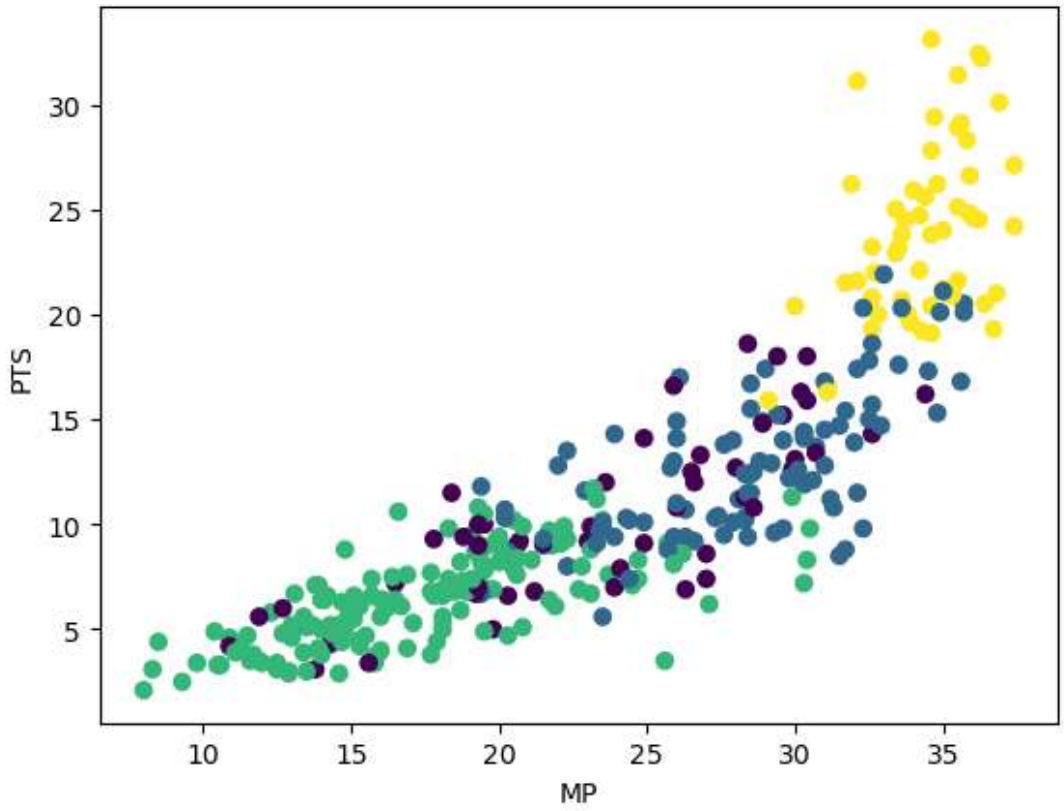


In [85]:
```
kmeans_model = KMeans(n_clusters = 4)

kmeans_model.fit(scaled_data)

nba_df2["clusters"] = kmeans_model.labels_

print(nba_df2.head())
print(nba_df2.shape)
```

```
               Player Pos  Age   Tm   G  GS    MP   FG   FGA    FG%
...  ORB  \
Rk
...
1   Precious Achiuwa   C   23  TOR  55  12  20.7  3.6   7.3  0.485
...  1.8
2       Steven Adams   C   29  MEM  42  42  27.0  3.7   6.3  0.597
...  5.1
3        Bam Adebayo   C   25  MIA  75  75  34.6  8.0  14.9  0.540
...  2.5
4       Ochai Agbaji  SG   22  UTA  59  22  20.5  2.8   6.5  0.427
...  0.7
5       Santi Aldama  PF   22  MEM  77  20  21.8  3.2   6.8  0.470
...  1.1

    DRB   TRB  AST  STL  BLK  TOV   PF   PTS  clusters
Rk
1   4.1   6.0  0.9  0.6  0.5  1.1  1.9   9.2         0
2   6.5  11.5  2.3  0.9  1.1  1.9  2.3   8.6         0
3   6.7   9.2  3.2  1.2  0.8  2.5  2.8  20.4         3
4   1.3   2.1  1.1  0.3  0.3  0.7  1.7   7.9         2
5   3.7   4.8  1.3  0.6  0.6  0.8  1.9   9.0         2

[5 rows x 30 columns]
(349, 30)

C:\Users\srken\AppData\Local\Temp\ipykernel_12912\2402600651.py:5: Set
tingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas
-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
(https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html
#returning-a-view-versus-a-copy)
  nba_df2["clusters"] = kmeans_model.labels_
```

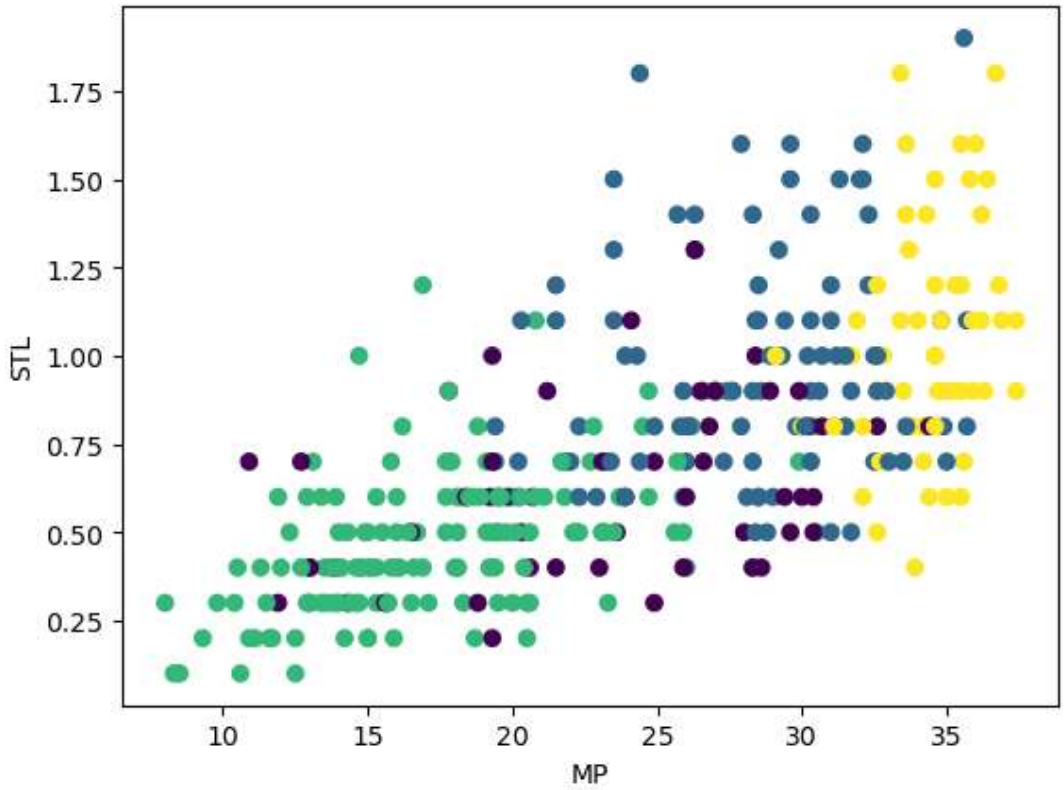
In [103]:
```
plt.scatter(nba_df2["MP"],
            nba_df2["PTS"],
            c = nba_df2["clusters"])
plt.xlabel("MP")
plt.ylabel("PTS")
```

Out[103]: Text(0, 0.5, 'PTS')



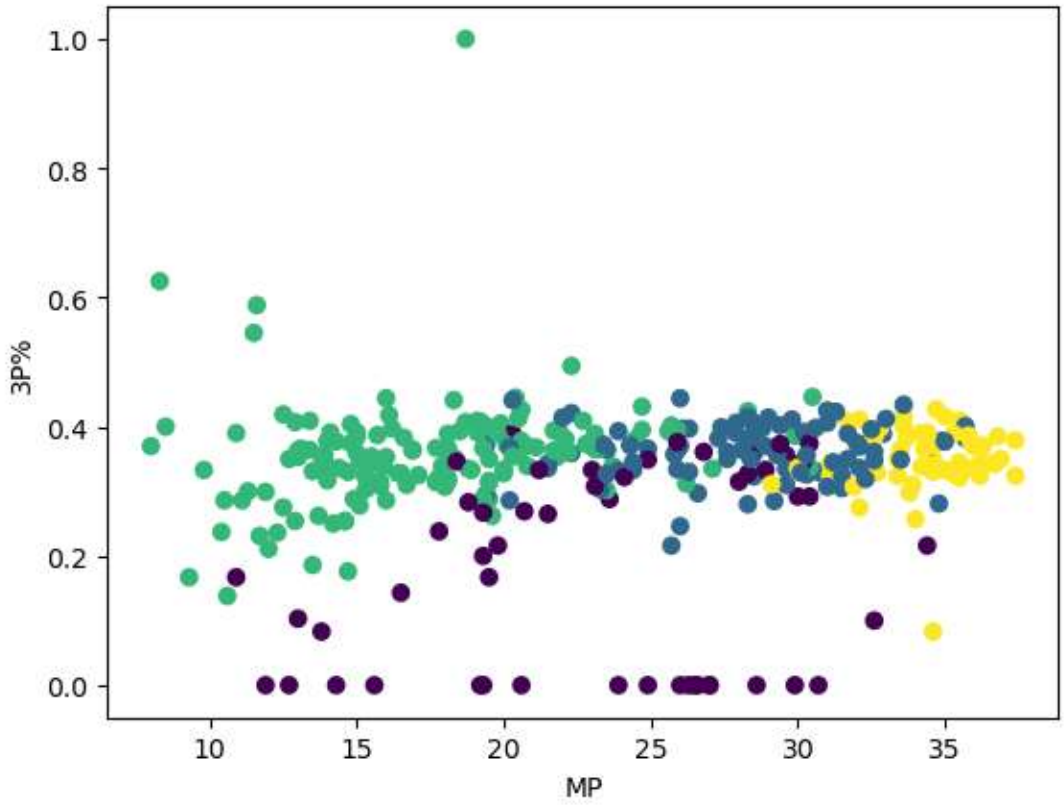
In [101]:
```
plt.scatter(nba_df2["MP"],
            nba_df2["STL"],
            c = nba_df2["clusters"])
plt.xlabel("MP")
plt.ylabel("STL")
```

Out[101]: Text(0, 0.5, 'STL')



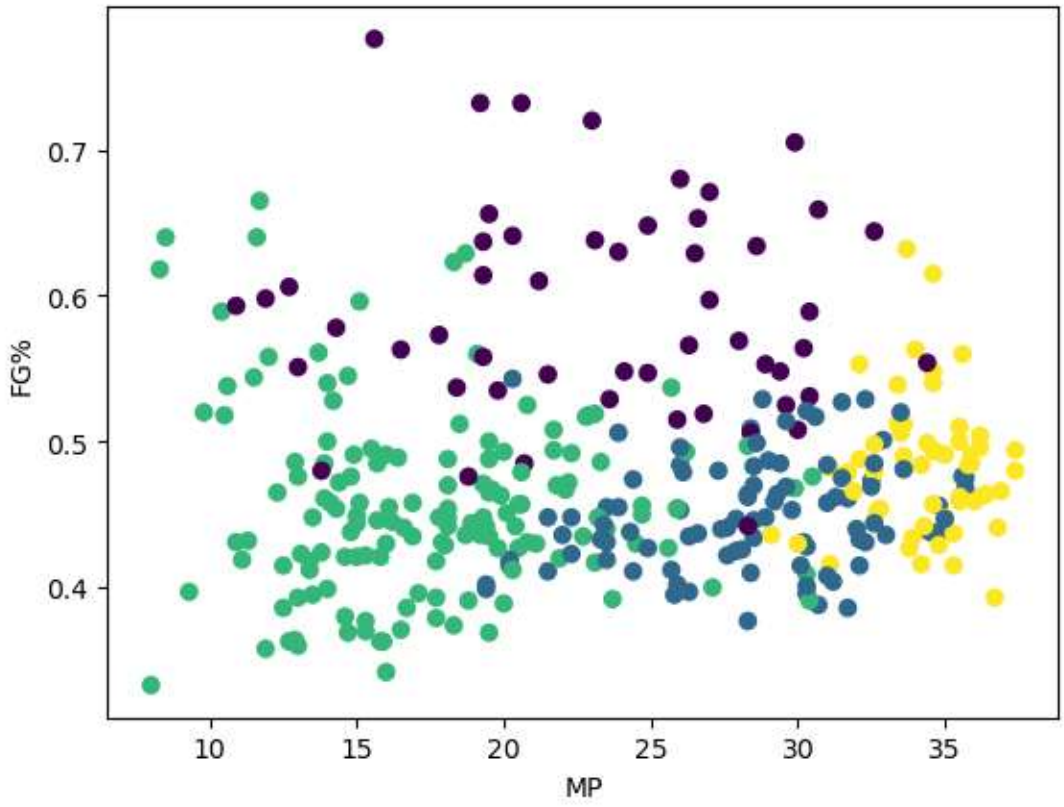
In [100]:
```
plt.scatter(nba_df2["MP"],
            nba_df2["3P%"],
            c = nba_df2["clusters"])
plt.xlabel("MP")
plt.ylabel("3P%")
```

Out[100]: Text(0, 0.5, '3P%')



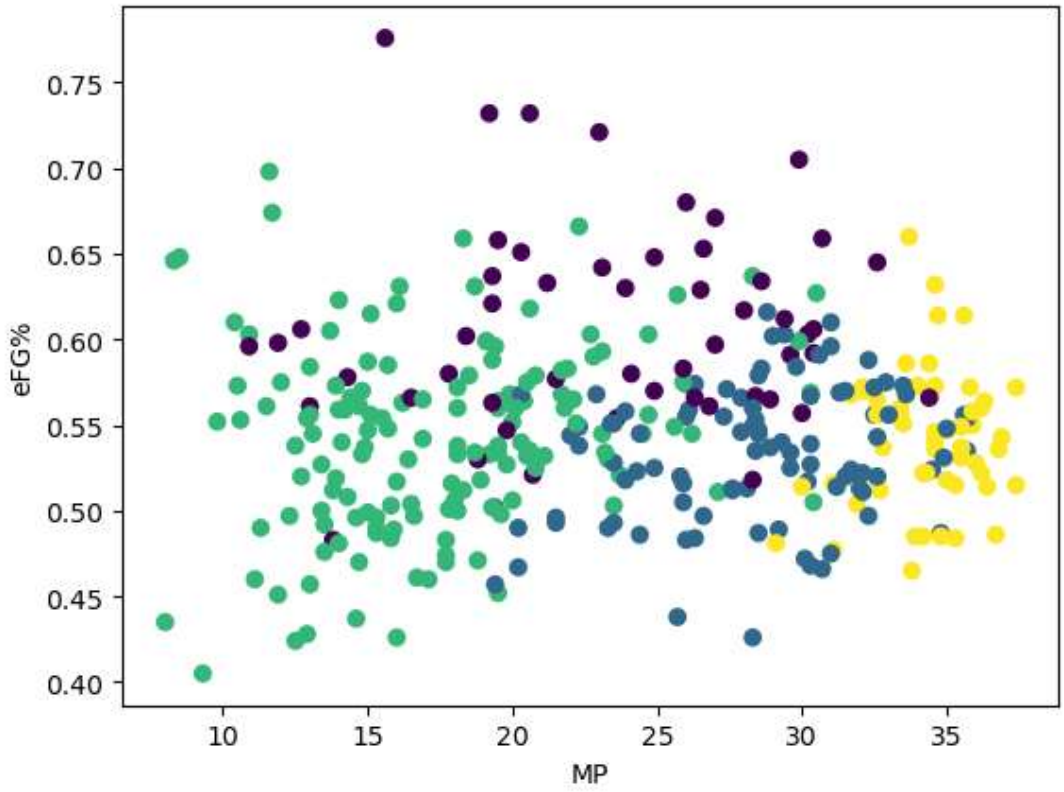
In [99]:
```
plt.scatter(nba_df2["MP"],
            nba_df2["FG%"],
            c = nba_df2["clusters"])
plt.xlabel("MP")
plt.ylabel("FG%")
```

Out[99]: Text(0, 0.5, 'FG%')



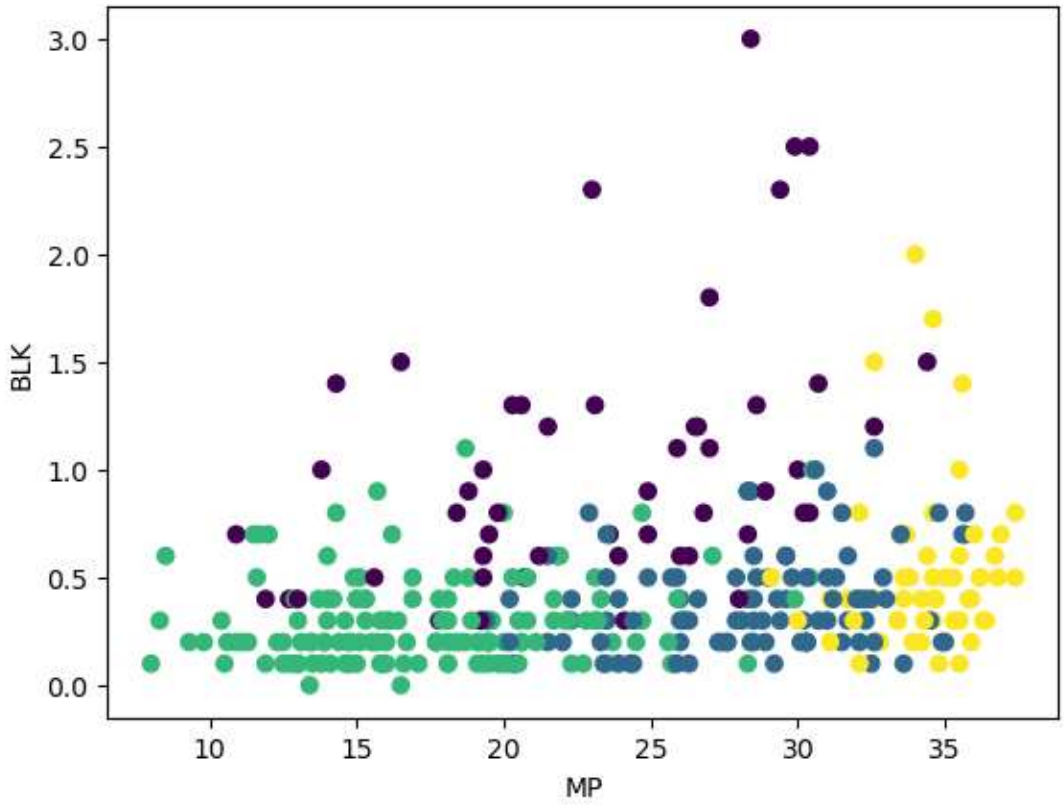
In [97]:
```
plt.scatter(nba_df2["MP"],
            nba_df2["eFG%"],
            c = nba_df2["clusters"])
plt.xlabel("MP")
plt.ylabel("eFG%")
```

Out[97]: Text(0, 0.5, 'eFG%')



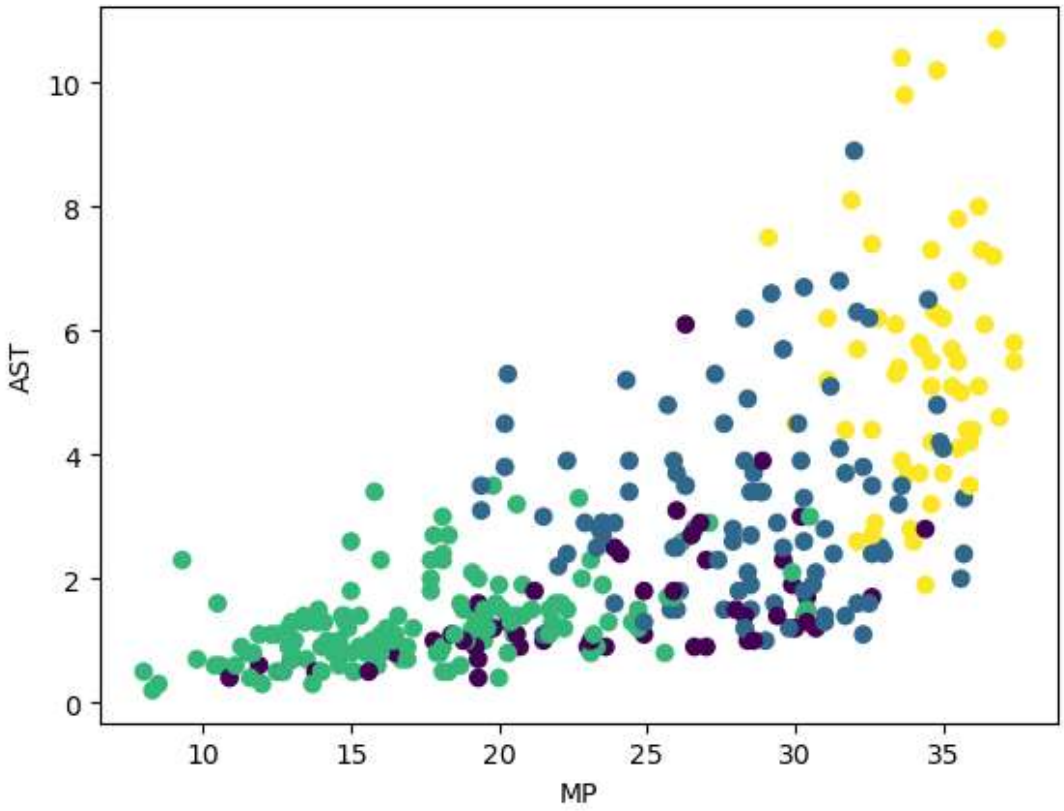
In [96]:
```
plt.scatter(nba_df2["MP"],
            nba_df2["BLK"],
            c = nba_df2["clusters"])
plt.xlabel("MP")
plt.ylabel("BLK")
```

Out[96]: Text(0, 0.5, 'BLK')



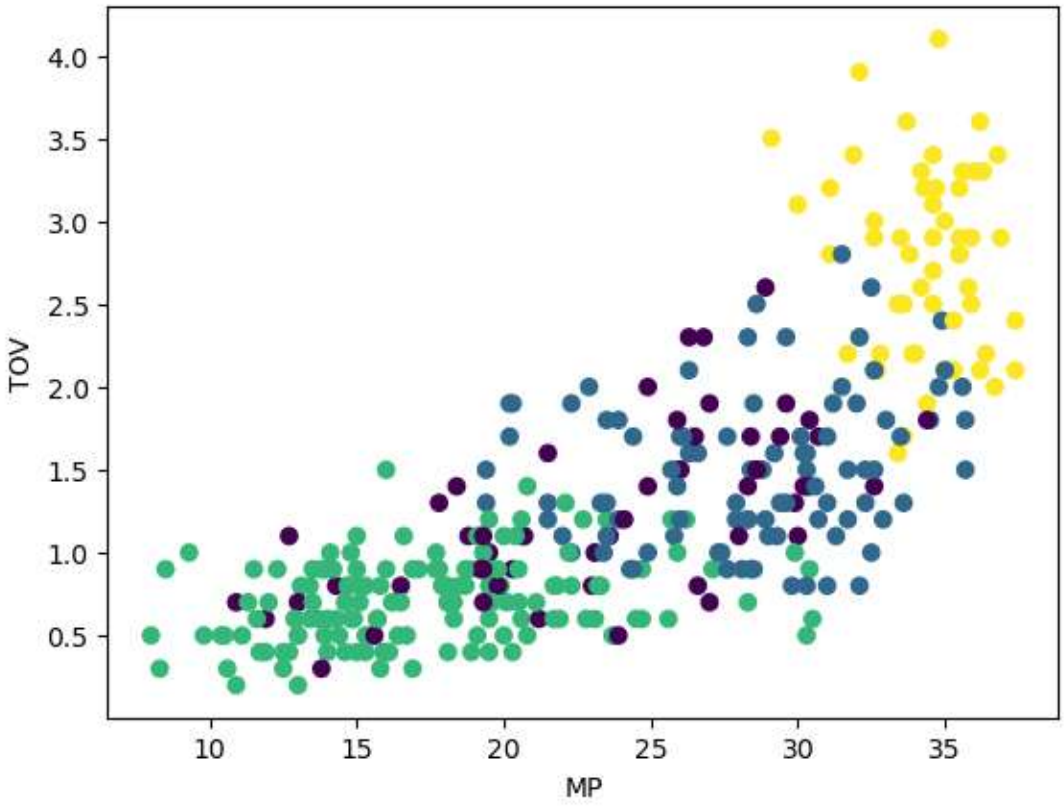
In [104]:
```
plt.scatter(nba_df2["MP"],
            nba_df2["AST"],
            c = nba_df2["clusters"])
plt.xlabel("MP")
plt.ylabel("AST")
```

Out[104]: Text(0, 0.5, 'AST')



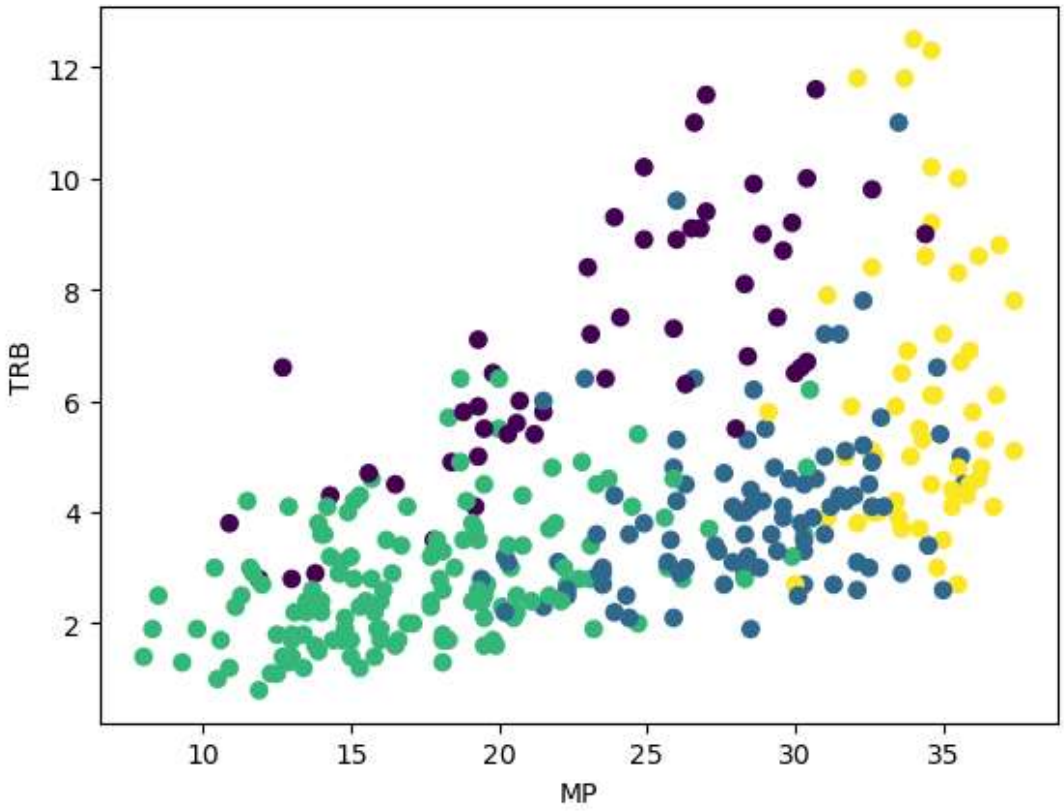
In [105]:
```
plt.scatter(nba_df2["MP"],
            nba_df2["TOV"],
            c = nba_df2["clusters"])
plt.xlabel("MP")
plt.ylabel("TOV")
```

Out[105]: Text(0, 0.5, 'TOV')



In [106]:
```
plt.scatter(nba_df2["MP"],
            nba_df2["TRB"],
            c = nba_df2["clusters"])
plt.xlabel("MP")
plt.ylabel("TRB")
```

Out[106]: Text(0, 0.5, 'TRB')



In [114]:
```
clust_2 = nba_df2[nba_df2['clusters'] == 2]
clust_2.head(10)
```

Out[114]:

Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%	2P
4	Ochai Agbaji	SG	22	UTA	59	22	20.5	2.8	6.5	0.427	1.4	3.9	0.355	1.4
5	Santi Aldama	PF	22	MEM	77	20	21.8	3.2	6.8	0.470	1.2	3.5	0.353	2.0
6	Nickeil Alexander-Walker	SG	24	TOT	59	3	15.0	2.2	5.0	0.444	1.0	2.7	0.384	1.2
22	Mo Bamba	C	24	TOT	49	7	15.7	2.4	4.9	0.485	1.0	2.5	0.387	1.4
30	Will Barton	SG	32	TOT	56	2	17.7	2.5	6.5	0.379	1.2	3.2	0.367	1.3
32	Keita Bates-Diop	SF	27	SAS	67	42	21.7	3.5	6.9	0.508	0.8	2.1	0.394	2.7
33	Nicolas Batum	PF	34	LAC	78	19	21.9	2.1	4.9	0.420	1.6	4.1	0.391	0.5
34	Darius Bazley	PF	22	TOT	43	1	14.3	1.9	4.3	0.454	0.5	1.2	0.377	1.5
37	MarJon Beauchamp	SF	22	MIL	52	11	13.5	1.8	4.7	0.395	0.9	2.7	0.331	0.9
38	D?vis Bert?ns	PF	30	DAL	45	1	10.9	1.5	3.6	0.431	1.2	3.1	0.390	0.3

In [111]:
```
clust_0 = nba_df2[nba_df2['clusters'] == 0]
clust_0.head(10)
```

Out[111]:

Tm	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	eFG%	FT	FTA	
TOR	55	12	20.7	3.6	7.3	0.485	0.5	2.0	0.269	3.0	5.4	0.564	0.521	1.6	2.3	
MEM	42	42	27.0	3.7	6.3	0.597	0.0	0.0	0.000	3.7	6.2	0.599	0.597	1.1	3.1	
CLE	68	68	32.6	5.9	9.2	0.644	0.0	0.1	0.100	5.9	9.1	0.653	0.645	2.4	3.3	
PHO	67	67	30.4	7.8	13.2	0.589	0.1	0.4	0.292	7.7	12.9	0.597	0.592	2.3	3.0	
DET	42	25	23.6	4.8	9.1	0.529	0.5	1.6	0.288	4.4	7.5	0.579	0.554	1.9	2.6	
PHO	61	14	14.3	2.0	3.4	0.578	0.0	0.0	0.000	2.0	3.4	0.578	0.578	0.4	1.1	
ORL	70	33	21.5	3.7	6.8	0.546	0.4	1.6	0.265	3.3	5.2	0.633	0.577	1.2	1.5	
ATL	65	63	26.6	5.4	8.2	0.653	0.0	0.0	0.000	5.4	8.2	0.654	0.653	1.2	2.0	
ORL	57	54	29.6	5.6	10.8	0.525	1.4	3.9	0.356	4.2	6.8	0.624	0.591	2.5	3.4	
MEM	56	8	19.5	4.1	6.2	0.656	0.0	0.1	0.167	4.1	6.1	0.665	0.658	1.8	2.4	

In [113]:
```
clust_1 = nba_df2[nba_df2['clusters'] == 1]
clust_1.head(10)
```

Out[113]:

Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2P
7	Grayson Allen	SG	27	MIL	72	70	27.4	3.4	7.7	0.440	2.0	5.1	0.399	1.4	2
9	Jose Alvarado	PG	24	NOP	61	10	21.5	3.3	8.0	0.411	1.4	4.0	0.336	1.9	4
10	Kyle Anderson	PF	29	MIN	69	46	28.4	3.7	7.2	0.509	0.6	1.5	0.410	3.0	5
13	Cole Anthony	PG	22	ORL	60	4	25.9	4.6	10.2	0.454	1.3	3.4	0.364	3.4	6
14	OG Anunoby	SF	25	TOR	67	67	35.6	6.3	13.2	0.476	2.1	5.5	0.387	4.2	7
16	Deni Avdija	SF	22	WAS	76	40	26.6	3.3	7.6	0.437	0.9	3.1	0.297	2.4	4
27	Harrison Barnes	PF	30	SAC	82	82	32.5	4.6	9.6	0.473	1.6	4.3	0.374	3.0	5
28	Scottie Barnes	PF	21	TOR	77	76	34.8	6.0	13.2	0.456	0.8	2.9	0.281	5.2	10
36	Malik Beasley	SG	26	TOT	81	27	25.8	4.6	11.6	0.395	2.9	8.1	0.357	1.7	3
40	Saddiq Bey	SF	23	TOT	77	37	27.6	4.6	10.9	0.422	2.0	5.4	0.361	2.6	5

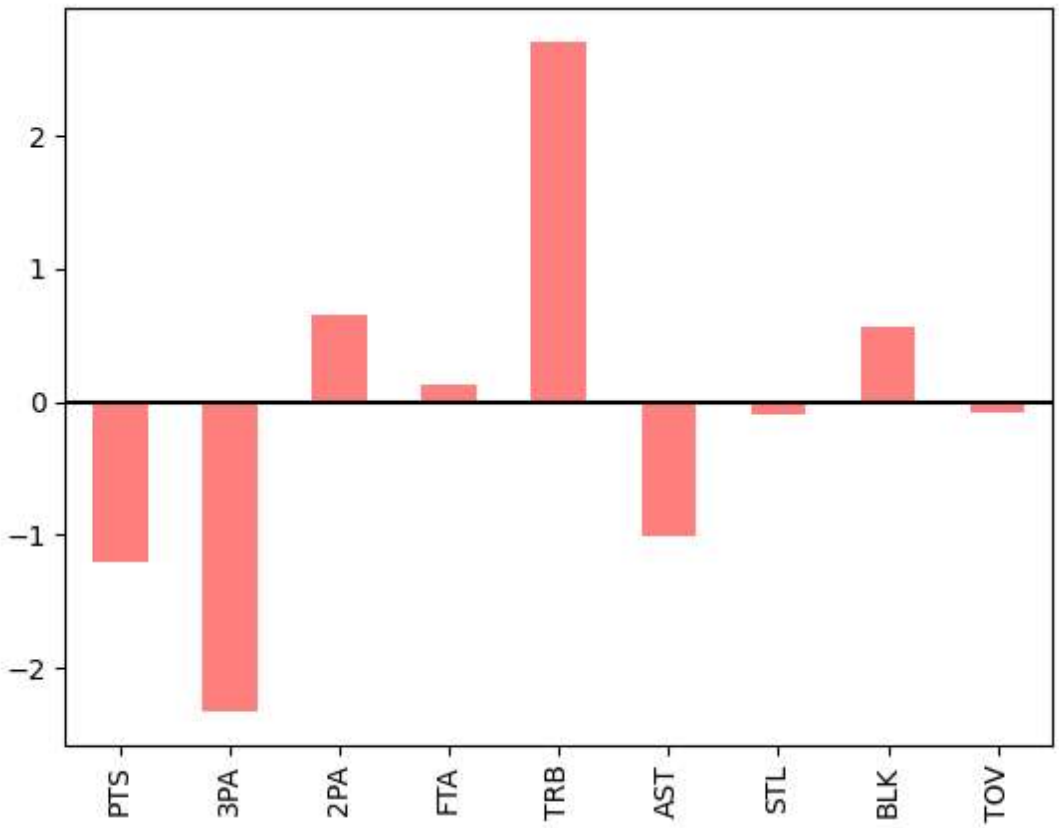
In [115]:
```
clust_3 = nba_df2[nba_df2['clusters'] == 3]
clust_3.head(10)
```

Out[115]:

Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	FG%	3P	3PA	3P%
3	Bam Adebayo	C	25	MIA	75	75	34.6	8.0	14.9	0.540	0.0	0.2	0.083
11	Giannis Antetokounmpo	PF	28	MIL	63	63	32.1	11.2	20.3	0.553	0.7	2.7	0.275
23	Paolo Banchero	PF	20	ORL	72	72	33.8	6.7	15.6	0.427	1.2	4.0	0.298
24	Desmond Bane	SG	24	MEM	58	58	31.7	7.8	16.2	0.479	2.9	7.0	0.408
29	RJ Barrett	SG	22	NYK	73	73	33.9	7.0	16.1	0.434	1.7	5.3	0.310
35	Bradley Beal	SG	29	WAS	50	50	33.5	8.9	17.6	0.506	1.6	4.4	0.365
46	Bojan Bogdanovi?	PF	33	DET	59	59	32.1	7.3	14.9	0.488	2.5	6.0	0.411
49	Devin Booker	SG	26	PHO	53	53	34.6	9.9	20.1	0.494	2.1	6.0	0.351
64	Jaylen Brown	SF	26	BOS	67	67	35.9	10.1	20.6	0.491	2.4	7.3	0.335
69	Jalen Brunson	PG	26	NYK	68	68	35.0	8.6	17.6	0.491	2.0	4.7	0.416

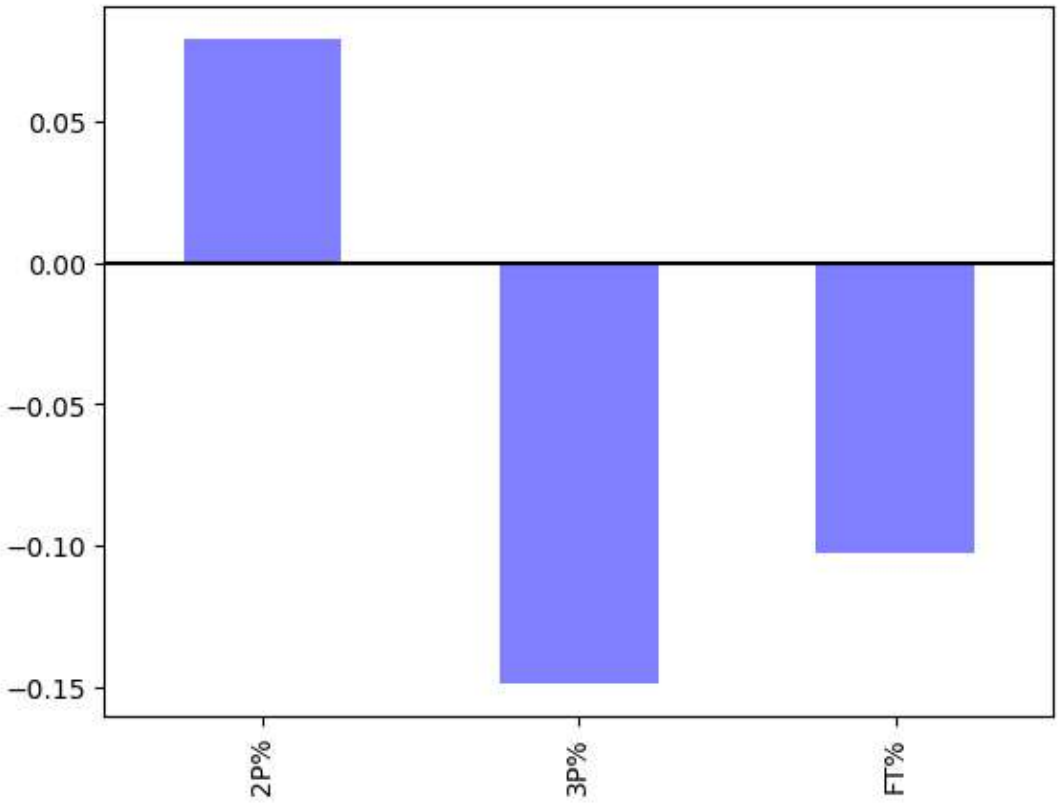
In [152]:
```
clust_0_count = (clust_0[["PTS", "3PA", "2PA",
                  "FTA", "TRB", "AST", "STL", "BLK", "TOV"]].mean() -
            nba_df2[["PTS", "3PA", "2PA",
                  "FTA", "TRB", "AST", "STL", "BLK", "TOV"]].mean())

clust_0_count.plot(kind = 'bar', color = 'r', alpha= 0.5)
plt.axhline(y = 0, color = 'black', linestyle = '-')
```

Out[152]: <matplotlib.lines.Line2D at 0x166f8f408e0>



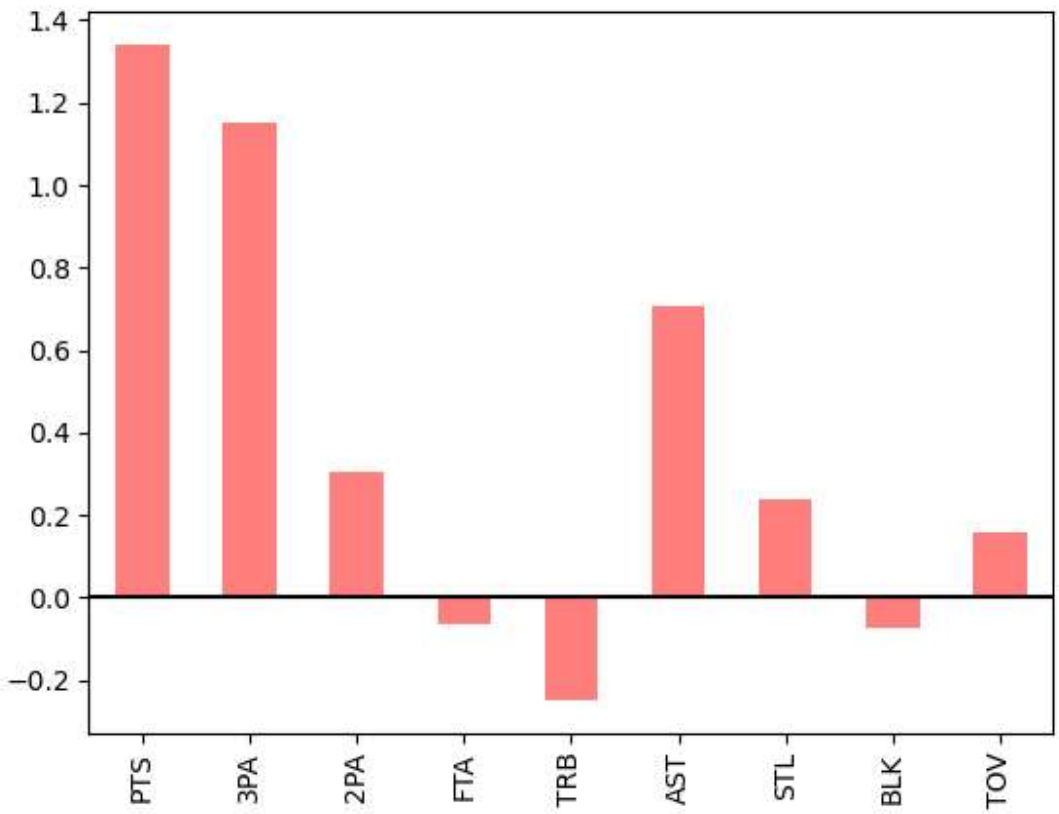
In [153]:
```
clust_0_pctg = (clust_0[["2P%", "3P%", "FT%"]].mean() -
           nba_df2[["2P%", "3P%", "FT%"]].mean())

clust_0_pctg.plot(kind = 'bar', color = 'b', alpha= 0.5)
plt.axhline(y = 0, color = 'black', linestyle = '-')
```

Out[153]: <matplotlib.lines.Line2D at 0x166fa1e9610>



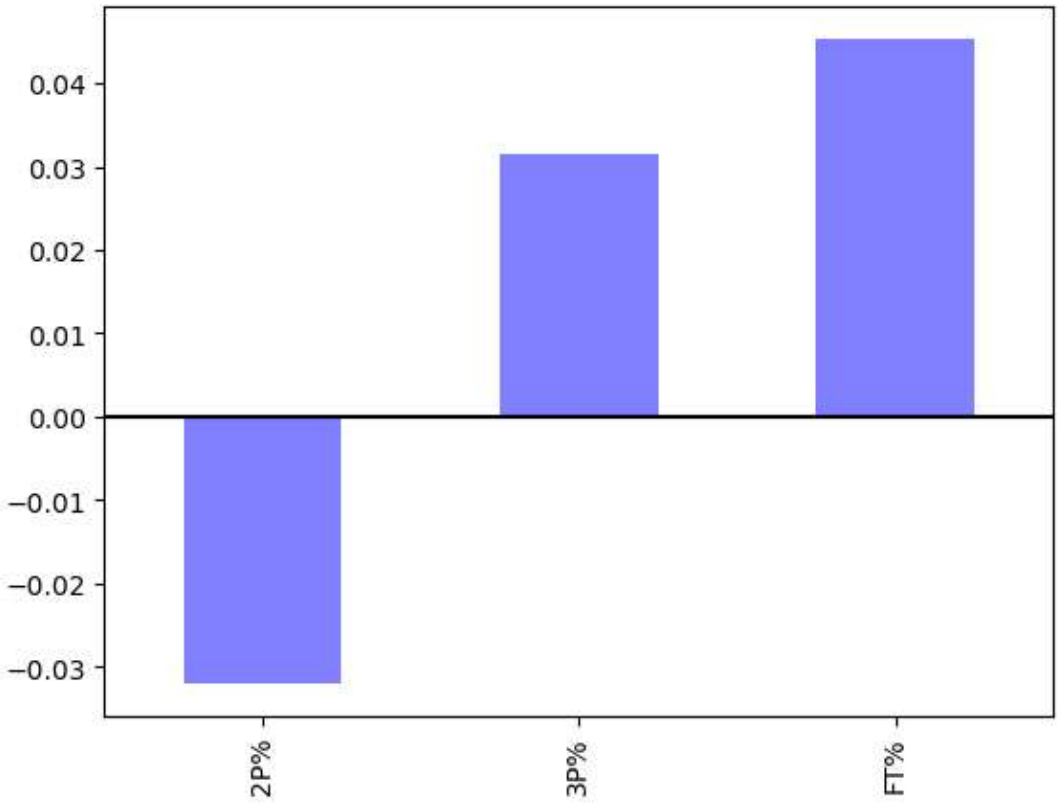
In [150]:
```
clust_1_count = (clust_1[["PTS", "3PA", "2PA",
                  "FTA", "TRB", "AST", "STL", "BLK", "TOV"]].mean() -
            nba_df2[["PTS", "3PA", "2PA",
                  "FTA", "TRB", "AST", "STL", "BLK", "TOV"]].mean())

clust_1_count.plot(kind = 'bar', color = 'r', alpha= 0.5)
plt.axhline(y = 0, color = 'black', linestyle = '-')
```

Out[150]: <matplotlib.lines.Line2D at 0x166f90d9d60>



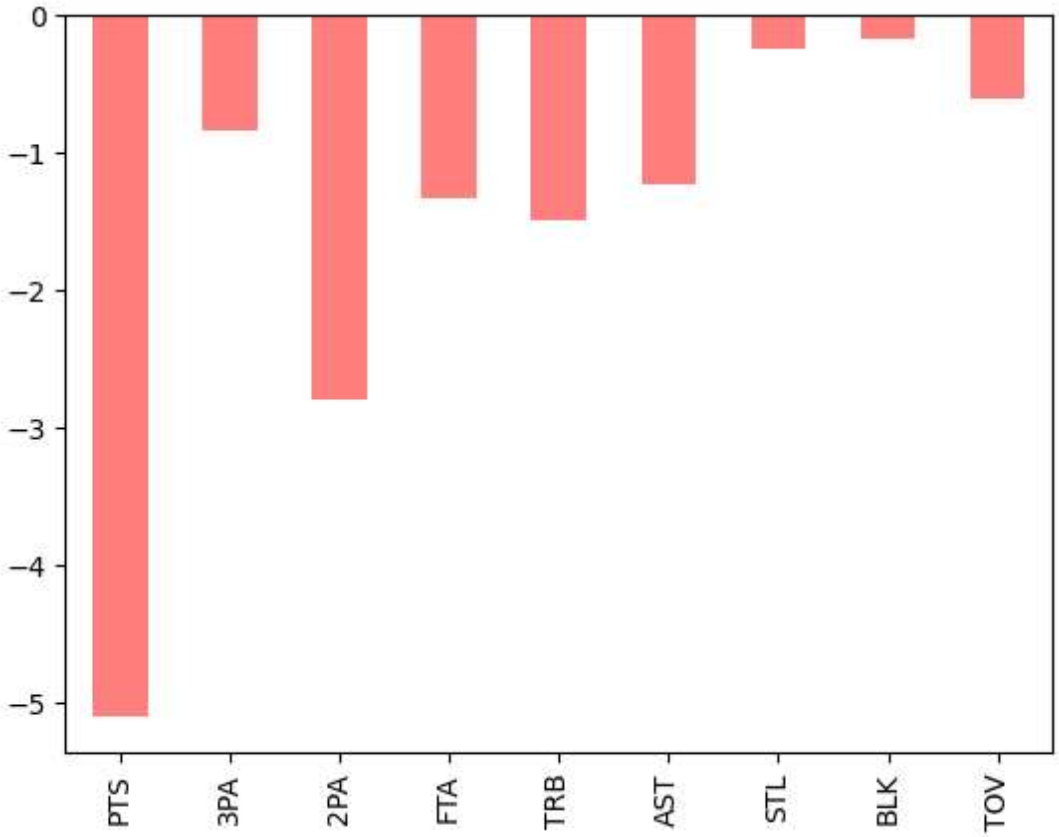
In [151]:
```
clust_1_pctg = (clust_1[["2P%", "3P%", "FT%"]].mean() -
           nba_df2[["2P%", "3P%", "FT%"]].mean())

clust_1_pctg.plot(kind = 'bar', color = 'b', alpha= 0.5)
plt.axhline(y = 0, color = 'black', linestyle = '-')
```

Out[151]: <matplotlib.lines.Line2D at 0x166f8fc9490>



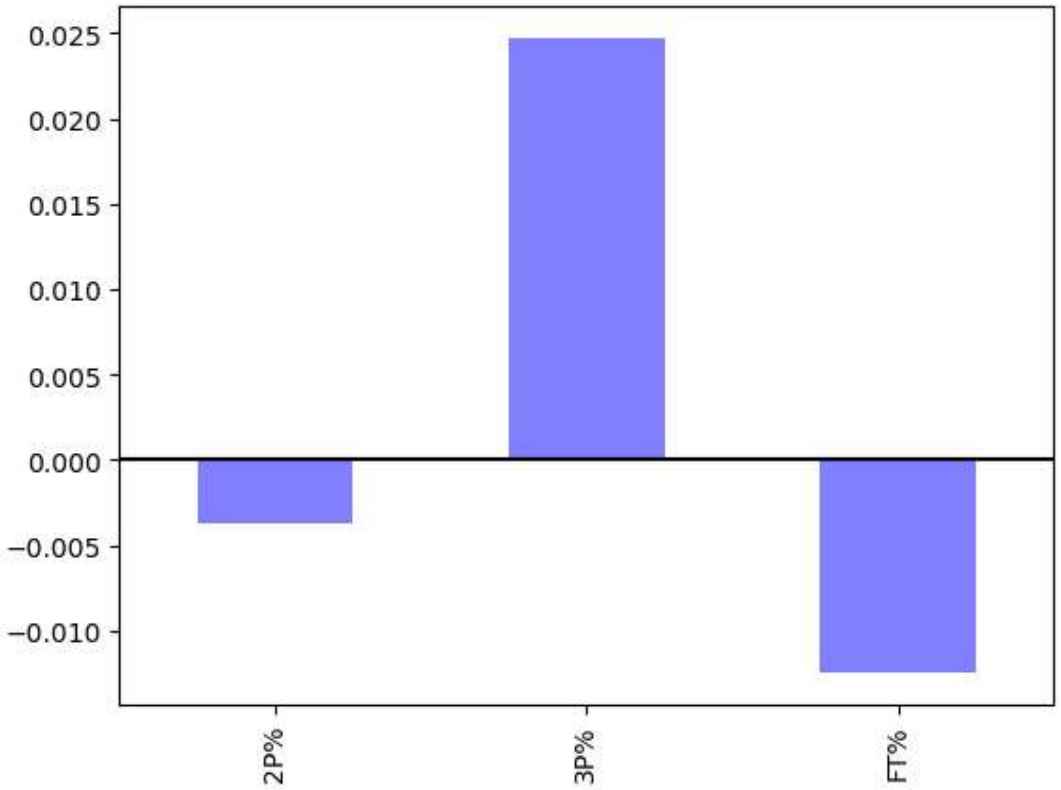
In [140]:
```
clust_2_count = (clust_2[["PTS", "3PA", "2PA",
                  "FTA", "TRB", "AST", "STL", "BLK", "TOV"]].mean() -
            nba_df2[["PTS", "3PA", "2PA",
                  "FTA", "TRB", "AST", "STL", "BLK", "TOV"]].mean())

clust_2_count.plot(kind = 'bar', color = 'r', alpha= 0.5)
```

Out[140]: <AxesSubplot:>



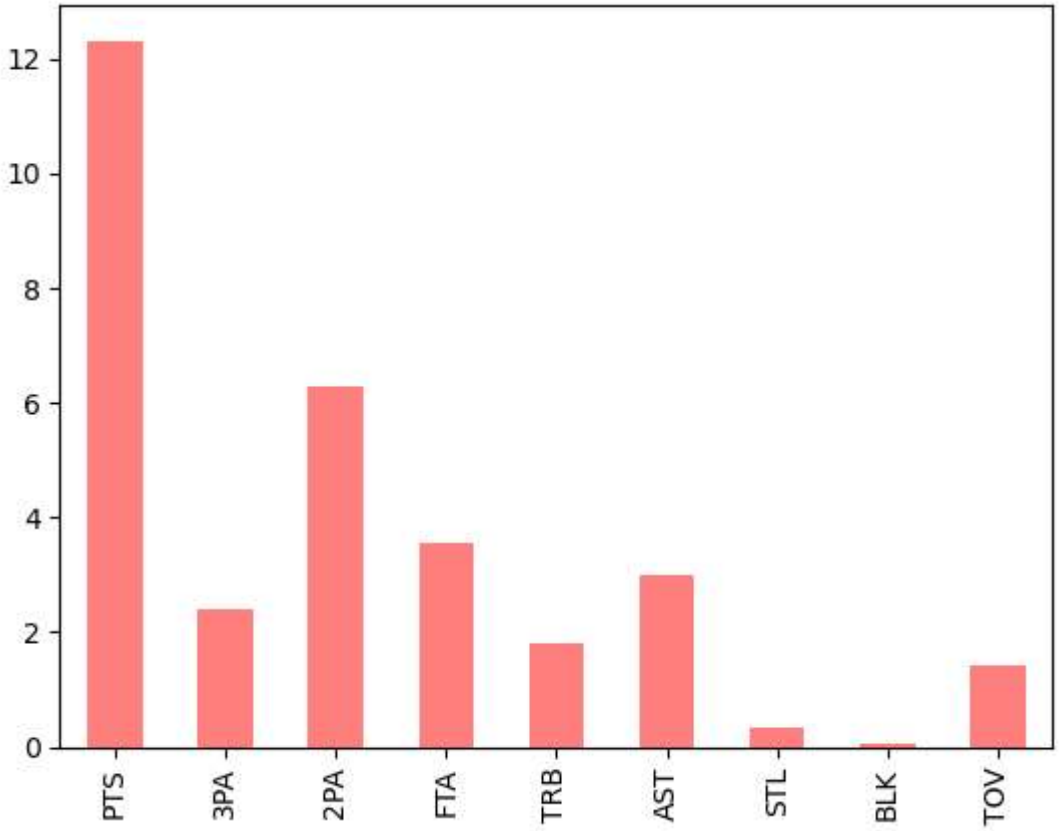
In [147]:
```
clust_2_pctg = (clust_2[["2P%", "3P%", "FT%"]].mean() -
           nba_df2[["2P%", "3P%", "FT%"]].mean())

clust_2_pctg.plot(kind = 'bar', color = 'b', alpha= 0.5)
plt.axhline(y = 0, color = 'black', linestyle = '-')
```

Out[147]: <matplotlib.lines.Line2D at 0x166f8f4af40>



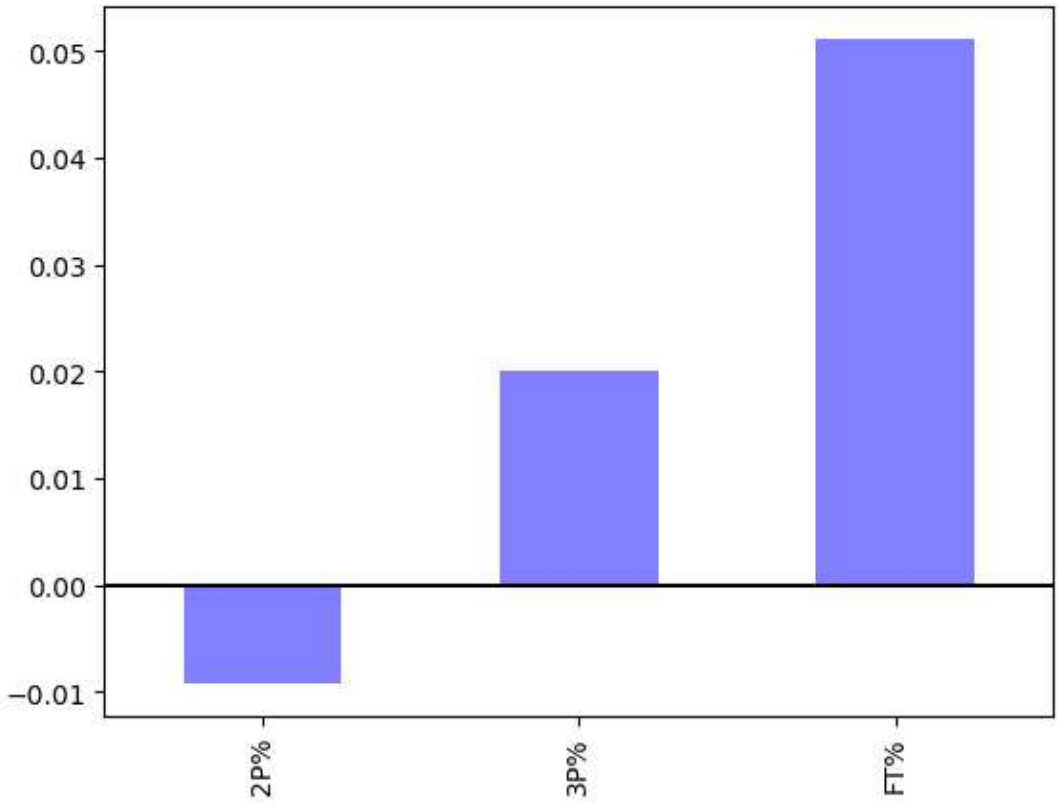
In [139]:
```
clust_3_count = (clust_3[["PTS", "3PA", "2PA",
                  "FTA", "TRB", "AST", "STL", "BLK", "TOV"]].mean() -
            nba_df2[["PTS", "3PA", "2PA",
                  "FTA", "TRB", "AST", "STL", "BLK", "TOV"]].mean())

clust_3_count.plot(kind = 'bar', color = 'r', alpha= 0.5)
```

Out[139]: <AxesSubplot:>



In [148]:
```
clust_3_pctg = (clust_3[["2P%", "3P%", "FT%"]].mean() -
           nba_df2[["2P%", "3P%", "FT%"]].mean())

clust_3_pctg.plot(kind = 'bar', color = 'b', alpha= 0.5)
plt.axhline(y = 0, color = 'black', linestyle = '-')
```

Out[148]: <matplotlib.lines.Line2D at 0x166f8fb8d90>



In [ ]:
```
```