

```
In [433]: import pandas as pd
import numpy as np
import requests
import warnings
from sklearn.model_selection import train_test_split, KFold
warnings.filterwarnings("ignore")
from sklearn.pipeline import Pipeline
from math import sqrt
```

```
In [436]: from sklearn.metrics import mean_squared_error
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import FunctionTransformer
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.pipeline import Pipeline
from sklearn.compose import ColumnTransformer

from sklearn.tree import DecisionTreeRegressor, DecisionTreeClassifier
from sklearn.model_selection import train_test_split, KFold
from sklearn.neighbors import KNeighborsRegressor
from sklearn.impute import SimpleImputer
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.linear_model import LogisticRegression, SGDClassifier
from sklearn.ensemble import VotingClassifier, ExtraTreesClassifier, RandomForestClassifier, GradientBoos
from sklearn.svm import SVC
```

```
In [3]: barts_hotspots = pd.read_csv('data/barts_hotspots.csv')
barts_to_all = pd.read_csv('data/barts_to_all.csv')
hotspots_to_all = pd.read_csv('data/hotspots_to_all.csv')
hours_q1 = pd.read_csv('data/hours_q1.csv')
hours_q2 = pd.read_csv('data/hours_q2.csv')
```

```
In [446]: barts_hotspots_filled.to_csv('barts_hotspots_updated.csv')
```

```
In [50]: barts_hotspots.columns = ['Date',  
    'Origin ID',  
    'Origin Name',  
    'Destination ID',  
    'Destination Name',  
    'Daily Mean Travel Time',  
    'Daily Range - Lower',  
    'Daily Range - Upper',  
    'AM Mean Travel Time',  
    'AM Range - Lower',  
    'AM Range - Upper',  
    'PM Mean Travel Time',  
    'PM Range - Lower Bound Travel Time',  
    'PM Range - Upper Bound Travel Time',  
    'Midday Mean Travel Time',  
    'Midday Range - Lower',  
    'Midday Range - Upper',  
    'Evening Mean Travel Time',  
    'Evening Range - Lower',  
    'Evening Range - Upper',  
    'Early Morning Mean Travel Time',  
    'Early Morning Range - Lower',  
    'Early Morning Range - Upper',  
    'weekday']
```

```
In [52]: barts_hotspots.sort_values('Date')
```

```
Out[52]:
```

	Date	Origin ID	Origin Name	Destination ID	Destination Name	Daily Mean Travel Time	Daily Range - Lower	Daily Range - Upper	AM Mean Travel Time	AM Range - Lower	...	Midday Mean Travel Time	Midday Range - Lower	Midday Range - Upper	Evening Mean Travel Time
<b>2106</b>	2019-01-01	3792	Oracle Park, 24 Willie Mays Plaza, San Francis...	3603	Embarcadero, San Francisco, CA	684.0	381.0	1228.0	NaN	NaN	...	432.0	338.0	551.0	NaN
<b>2524</b>	2019-01-01	3792	Oracle Park, 24 Willie Mays Plaza, San Francis...	3692	2nd Street and Stevenson Street (Montgomery BA...	483.0	319.0	729.0	NaN	NaN	...	464.0	343.0	627.0	427.0
<b>2091</b>	2019-01-01	3603	Embarcadero, San Francisco, CA	3396	The Palace Of Fine Arts, 3601 Lyon St, San Fra...	1315.0	1074.0	1609.0	NaN	NaN	...	NaN	NaN	NaN	NaN

```
In [64]: def stats(df):
        stats = df.describe().T
        stats['filled'] = stats['count']/len(df)
        return stats[['filled']]
```

```
In [160]: arts_hotspots_filled = barts_hotspots.copy()
arts_hotspots_filled['Destination Name'] = barts_hotspots_filled['Destination Name'].apply(lambda x: "Fisherman's Wharf")
arts_hotspots_filled['Destination Name'].nunique()
arts_hotspots_filled['Origin Name'] = barts_hotspots_filled['Origin Name'].apply(lambda x: "Fisherman's Wharf")
arts_hotspots_filled['Origin Name'].nunique()
```

```
Out[160]: 6
```

```
In [85]: def additional_fill(df,col):
        filled = df.copy()
        if type(col) == 'list':
            for i in col:
                df_filled[i] = barts_hotspots_filled.groupby(['Origin ID','weekday','Destination ID'])[i].transform(lambda x: x.fillna(0))
        return df_filled
```

```
In [93]: col_to_fill = list(barts_hotspots_filled.columns)
col_to_fill.remove('Date')
col_to_fill.remove('Origin ID')
col_to_fill.remove('Origin Name')
col_to_fill.remove('Destination ID')
col_to_fill.remove('Destination Name')
col_to_fill
```

```
Out[93]: ['Daily Mean Travel Time',
'Daily Range - Lower',
'Daily Range - Upper',
'AM Mean Travel Time',
'AM Range - Lower',
'AM Range - Upper',
'PM Mean Travel Time',
'PM Range - Lower Bound Travel Time',
'PM Range - Upper Bound Travel Time',
'Midday Mean Travel Time',
'Midday Range - Lower',
'Midday Range - Upper',
'Evening Mean Travel Time',
'Evening Range - Lower',
'Evening Range - Upper',
'Early Morning Mean Travel Time',
'Early Morning Range - Lower',
'Early Morning Range - Upper',
'weekday']
```

```
In [94]: barts_hotspots_filled = conditional_fill(barts_hotspots_filled,col_to_fill)
```

```
In [157]: stats(barts_hotspots_filled)
```

```
Out[157]:
```

	filled
<b>Origin ID</b>	1.000000
<b>Destination ID</b>	1.000000
<b>Daily Mean Travel Time</b>	1.000000
<b>Daily Range - Lower</b>	1.000000
<b>Daily Range - Upper</b>	1.000000
<b>AM Mean Travel Time</b>	0.857274
<b>AM Range - Lower</b>	0.857274
<b>AM Range - Upper</b>	0.857274
<b>PM Mean Travel Time</b>	0.960098
<b>PM Range - Lower Bound Travel Time</b>	0.960098
<b>PM Range - Upper Bound Travel Time</b>	0.960098
<b>Midday Mean Travel Time</b>	1.000000
<b>Midday Range - Lower</b>	1.000000
<b>Midday Range - Upper</b>	1.000000
<b>Evening Mean Travel Time</b>	0.921117
<b>Evening Range - Lower</b>	0.921117
<b>Evening Range - Upper</b>	0.921117
<b>Early Morning Mean Travel Time</b>	0.588091
<b>Early Morning Range - Lower</b>	0.588091
<b>Early Morning Range - Upper</b>	0.588091
<b>weekday</b>	1.000000

```
In [96]: stats(barts_hotspots)
```

```
Out[96]:
```

	filled
<b>Origin ID</b>	1.000000
<b>Destination ID</b>	1.000000
<b>Daily Mean Travel Time</b>	0.982198
<b>Daily Range - Lower</b>	0.982198
<b>Daily Range - Upper</b>	0.982198
<b>AM Mean Travel Time</b>	0.550031
<b>AM Range - Lower</b>	0.550031
<b>AM Range - Upper</b>	0.550031
<b>PM Mean Travel Time</b>	0.715163
<b>PM Range - Lower Bound Travel Time</b>	0.715163
<b>PM Range - Upper Bound Travel Time</b>	0.715163
<b>Midday Mean Travel Time</b>	0.776857
<b>Midday Range - Lower</b>	0.776857
<b>Midday Range - Upper</b>	0.776857
<b>Evening Mean Travel Time</b>	0.685083
<b>Evening Range - Lower</b>	0.685083
<b>Evening Range - Upper</b>	0.685083
<b>Early Morning Mean Travel Time</b>	0.259975
<b>Early Morning Range - Lower</b>	0.259975
<b>Early Morning Range - Upper</b>	0.259975
<b>weekday</b>	1.000000

```
In [109]: barts_hotspots_filled.head()
```

```
Out[109]:
```

Origin Name	Daily Mean Travel Time	Daily Range - Lower	Daily Range - Upper	AM Mean Travel Time	AM Range - Lower	AM Range - Upper	...	Midday Mean Travel Time	Midday Range - Lower	Midday Range - Upper	Evening Mean Travel Time	Evening Range - Lower	Evening Range - Upper	Early Morning Mean Travel Time	Early Morning Range - Lower	Early Morning Range - Upper	weekday
lero, San , CA	1588.0	1054.0	2392.0	NaN	NaN	NaN	...	1398.5	1191.5	1640.5	1149.5	861.0	1535.0	NaN	NaN	NaN	6
an's San , CA	639.0	512.0	796.0	579.0	427.0	731.0	...	654.0	524.0	817.0	608.0	492.0	753.0	NaN	NaN	NaN	6
an's San , CA	730.0	520.0	1024.0	567.0	460.0	674.0	...	791.0	590.0	1061.0	569.0	406.0	797.0	NaN	NaN	NaN	5
lero, San , CA	512.0	371.0	705.0	648.0	461.0	835.0	...	528.0	398.0	702.0	442.0	321.0	608.0	363.0	297.0	443.0	3
lero, San , CA	505.0	364.0	701.0	621.0	489.0	753.0	...	429.0	339.0	542.0	502.0	354.0	711.0	377.0	311.0	467.0	0

```
In [475]: ', "Origin Name", 'Date', 'weekday', 'Midday Mean Travel Time', 'Evening Mean Travel Time', 'Early Morning Mean
```

```

In [476]: num = 200
test_barts_hotspots_filled_melt = pd.melt(test_barts_hotspots_filled, id_vars=['Destination Name', 'Origin Name', 'Date', 'weekday', 'time_of_day'],
var_name="time_of_day",
value_name="Value")
test_barts_hotspots_filled_melt_sample = test_barts_hotspots_filled_melt.sample(num)
test_barts_hotspots_filled_melt_sample.head(num)

```

Out[476]:

	Destination Name	Origin Name	Date	weekday	time_of_day	Value
<b>8506</b>	Embarcadero, San Francisco, CA	The Palace Of Fine Arts, 3601 Lyon St, San Francisco, CA	2019-02-25	0	Early Morning Mean Travel Time	739.5
<b>297</b>	Embarcadero, San Francisco, CA	Fisherman's Wharf, 286-298 Jefferson St, San Francisco, CA	2019-04-26	4	Midday Mean Travel Time	812.0
<b>3127</b>	Fisherman's Wharf, 286-298 Jefferson St, San Francisco, CA	Powell BART Station, Market St and Powell St, San Francisco, CA	2019-03-01	4	Midday Mean Travel Time	873.0
<b>5988</b>	The Palace Of Fine Arts, 3601 Lyon St, San Francisco, CA	Powell BART Station, Market St and Powell St, San Francisco, CA	2019-03-26	1	Evening Mean Travel Time	1118.0
<b>3926</b>	2nd Street and Stevenson Street, San Francisco, CA	Oracle Park, 24 Willie Mays Plaza, San Francisco, CA	2019-04-03	2	Evening Mean Travel Time	449.0
<b>4489</b>	Powell BART Station, Market St and Powell St, San Francisco, CA	Oracle Park, 24 Willie Mays Plaza, San Francisco, CA	2019-04-06	5	Evening Mean Travel Time	626.0
<b>4117</b>	Fisherman's Wharf, 286-298 Jefferson St, San Francisco, CA	2nd Street and Stevenson Street, San Francisco, CA	2019-05-27	0	Evening Mean Travel Time	571.0



```
In [255]: def recommend(df, origin, destination, **argv):
          copy = df.copy()
          origin_filter = copy['Origin Name'].apply(lambda x: True if origin in x else False)
          destination_filter = copy['Destination Name'].apply(lambda x: True if destination in x else False)
          copy = copy[origin_filter&destination_filter]
          if not argv:
              return 'no arg'
          for key,value in argv.items():
              temp_filter = copy[key].apply(lambda x: True if str(value) in str(x) else False)
              copy = copy[temp_filter]
          output = copy.groupby('time_of_day')['Value'].mean().sort_values().index[0]
          return output

          recommend(test_barts_hotspots_filled_melt_sample, 'Powell' , 'Fisherman', weekday=2)
```

Out[255]: 'Early Morning Mean Travel Time'

```
In [477]: s_filled_melt.groupby(['Destination Name','Origin Name','weekday','time_of_day'])['Value'].mean().to_json
```

In [ ]:

In [ ]:

In [ ]:

```
In [470]: test_df = pd.DataFrame([[ 'startN','end1','week1','A',5],[ 'start2','end1','week1','A',3],
    [ 'start2','end1','week1','A',1],
    [ 'start3','end1','week2','P',2],
    [ 'start2','end1','week1','P',3],
    [ 'qqq','end2','week2','A',3],
    [ 'start1','end1','week1','A',6],
    [ 'start2','end1','week1','P',3],
    [ 'start2','end2','week2','A',3],
    [ 'start1','end1','week1','A',6],
    [ 'start2','end2','week2','A',3],
    [ 'start1','end1','week1','A',2],
    [ 'qqq','end1','week2','P',22],
    [ 'start2','end2','week2','A',0],
    [ 'start3','end1','week1','P',3],
    [ 'start3','end2','week2','P',3],
    [ 'start1','end2','week2','P',3],
    [ 'start3','end2','week2','A',99],
    [ 'start1','end1','week1','P',2],
    [ 'start3','end1','week2','P',22],
    [ 'start2','end2','week2','A',0],
    [ 'qqq','end1','week1','P',6],
    [ 'start2','end1','week1','P',0],
    [ 'start1','end1','week2','A',3],
    [ 'start1','end2','week1','P',6],
    [ 'start3','end2','week1','A',9]])

#test_df
```

```
In [469]: #pd.DataFrame(test_df.groupby([0,1,2,3])[4].apply(lambda x: x.mean())).sort_values(by=[0,1,4])
```

```
In [478]: test_df = pd.DataFrame(test_barts_hotspots_filled_melt.groupby(['Origin Name','Destination Name','weekda
```

```
In [466]: test_df.sort_values('Value').groupby(['Origin Name', 'Destination Name'])['Value'].transform('min')
```

```
Out[466]: 63      435.876349
          75      435.876349
          69      435.876349
          78      435.876349
          72      435.876349
          82      435.876349
          66      435.876349
          79      435.876349
          39      508.074798
          42      508.764161
          64      435.876349
          81      435.876349
          21      508.074798
          76      435.876349
          67      435.876349
          48      508.764161
          57      508.764161
          70      435.876349
          51      508.764161
          33      508.074798
          54      508.764161
          73      435.876349
          60      508.764161
          65      435.876349
          83      435.876349
          30      508.074798
          36      508.074798
          68      435.876349
          27      508.074798
          80      435.876349
          ...
           2      661.801843
          86      646.776286
          97      646.776286
          89      646.776286
           5      661.801843
          17      661.801843
          124     729.861483
          95      646.776286
          104     646.776286
           8      661.801843
```

14	661.801843
101	646.776286
11	661.801843
109	729.861483
92	646.776286
106	729.861483
100	646.776286
118	729.861483
98	646.776286
112	729.861483
115	729.861483
116	729.861483
87	646.776286
121	729.861483
107	729.861483
125	729.861483
110	729.861483
119	729.861483
113	729.861483
122	729.861483

Name: Value, Length: 126, dtype: float64

```
In [479]: output = test_df.sort_values('Value').drop_duplicates(['Origin Name', 'Destination Name'])
output
```

Out[479]:

	Origin Name	Destination Name	weekday	time_of_day	Value
228	Oracle Park, 24 Willie Mays Plaza, San Francis...	Embarcadero, San Francisco, CA	6	Early Morning Mean Travel Time	316.261758
189	Oracle Park, 24 Willie Mays Plaza, San Francis...	2nd Street and Stevenson Street, San Francisco...	0	Early Morning Mean Travel Time	347.520000
24	2nd Street and Stevenson Street, San Francisco...	Oracle Park, 24 Willie Mays Plaza, San Francis...	1	Early Morning Mean Travel Time	364.923077
99	Embarcadero, San Francisco, CA	Oracle Park, 24 Willie Mays Plaza, San Francis...	5	Early Morning Mean Travel Time	407.000000
81	Embarcadero, San Francisco, CA	Fisherman's Wharf, 286-298 Jefferson St, San F...	6	Early Morning Mean Travel Time	419.440926
165	Fisherman's Wharf, 286-298 Jefferson St, San F...	Embarcadero, San Francisco, CA	6	Early Morning Mean Travel Time	451.757366
273	Powell BART Station, Market St and Powell St, ...	Oracle Park, 24 Willie Mays Plaza, San Francis...	0	Early Morning Mean Travel Time	467.889048
249	Oracle Park, 24 Willie Mays Plaza, San Francis...	Powell BART Station, Market St and Powell St, ...	6	Early Morning Mean Travel Time	492.370921
0	2nd Street and Stevenson Street, San Francisco...	Fisherman's Wharf, 286-298 Jefferson St, San F...	0	Early Morning Mean Travel Time	493.440000
252	Powell BART Station, Market St and Powell St, ...	Fisherman's Wharf, 286-298 Jefferson St, San F...	0	Early Morning Mean Travel Time	525.852483
168	Fisherman's Wharf, 286-298 Jefferson St, San F...	Powell BART Station, Market St and Powell St, ...	0	Early Morning Mean Travel Time	602.264352
135	Fisherman's Wharf, 286-298 Jefferson St, San F...	2nd Street and Stevenson Street, San Francisco...	3	Early Morning Mean Travel Time	673.307692
120	Embarcadero, San Francisco, CA	The Palace Of Fine Arts, 3601 Lyon St, San Fra...	5	Early Morning Mean Travel Time	693.873412
60	2nd Street and Stevenson Street, San Francisco...	The Palace Of Fine Arts, 3601 Lyon St, San Fra...	6	Early Morning Mean Travel Time	729.707407
354	The Palace Of Fine Arts, 3601 Lyon St, San Fra...	Embarcadero, San Francisco, CA	6	Early Morning Mean Travel Time	756.205270
309	Powell BART Station, Market St and Powell St, ...	The Palace Of Fine Arts, 3601 Lyon St, San Fra...	5	Early Morning Mean Travel Time	762.163532

	Origin Name	Destination Name	weekday	time_of_day	Value
357	The Palace Of Fine Arts, 3601 Lyon St, San Fra...	Powell BART Station, Market St and Powell St, ...	0	Early Morning Mean Travel Time	839.975724
333	The Palace Of Fine Arts, 3601 Lyon St, San Fra...	2nd Street and Stevenson Street, San Francisco...	6	Early Morning Mean Travel Time	867.405528

```
In [480]: import json
temp = output.groupby('Origin Name')[['Destination Name', 'weekday', 'time_of_day', 'Value']].apply(lambda
with open('temp.json', 'w') as f:
    json.dump(temp, f, indent=4)
```

```
In [380]: test_barts_hotspots_filled.columns = ['Destination_Name',
'Origin_Name',
'Date',
'weekday',
'Midday Mean Travel Time',
'Evening Mean Travel Time',
'Early Morning Mean Travel Time']
```

```
In [396]: from collections import defaultdict
```

```
d = defaultdict(dict)
```

```
for row in test_barts_hotspots_filled.itertuples(index=False):
    d[row.Destination_Name][row.Origin_Name] = row.Date
```

```
pprint.pprint(d)
```

```
defaultdict(<class 'dict'>,
            {'2nd Street and Stevenson Street (Montgomery BART, San Francisco, CA': {"Fisherman's Wharf, 286-298 Jefferson St, San Francisco, CA": Timestamp('2019-03-13 00:00:00'),
                                                                                          'Oracle Park, 24 Willie Mays Plaza, San Francisco, CA': Timestamp('2019-01-02 00:00:00'),
                                                                                          'The Palace Of Fine Arts, 3601 Lyon St, San Francisco, CA': Timestamp('2019-03-01 00:00:00')},
            'Embarcadero, San Francisco, CA': {"Fisherman's Wharf, 286-298 Jefferson St, San Francisco, CA": Timestamp('2019-02-24 00:00:00'),
                                                                                          'Oracle Park, 24 Willie Mays Plaza, San Francisco, CA': Timestamp('2019-02-09 00:00:00'),
                                                                                          'The Palace Of Fine Arts, 3601 Lyon St, San Francisco, CA': Timestamp('2019-02-10 00:00:00')},
            "Fisherman's Wharf, 286-298 Jefferson St, San Francisco, CA": {"2nd Street and Stevenson Street, San Francisco, CA': Timestamp('2019-02-14 00:00:00'),
                                                                                          'Embarcadero, San Francisco, CA': Timestamp('2019-01-16 00:00:00'),
                                                                                          'Powell BART Station, Market St and Powell St, San Francisco, CA': Timestamp('2019-03-04 00:00:00')},
            'Oracle Park, 24 Willie Mays Plaza, San Francisco, CA': {"2nd Street and Stevenson Street, San Francisco, CA': Timestamp('2019-03-14 00:00:00'),
                                                                                          'Embarcadero, San Francisco, CA': Timestamp('2019-02-24 00:00:00'),
                                                                                          'Powell BART Station, Market St and Powell St, San Francisco, CA': Timestamp('2019-01-09 00:00:00')},
            'Powell BART Station, Market St and Powell St, San Francisco, CA': {"Fisherman's Wharf, 286-298 Jefferson St, San Francisco, CA": Timestamp('2019-03-13 00:00:00'),
                                                                                          'Oracle Park, 24 Willie Mays Plaza, San Francisco, CA': Timestamp('2019-03-20 00:00:00'),
                                                                                          'The Palace Of Fine Arts, 3601 Lyon St, San Francisco, CA': Timestamp('2019-01-24 00:00:00')},
            'The Palace Of Fine Arts, 3601 Lyon St, San Francisco, CA': {"2nd Street and Stevenson Street, San Francisco, CA': Timestamp('2019-03-31 00:00:00'),
                                                                                          'Embarcadero, San Francisco,
```

```
CA': Timestamp('2019-03-03 00:00:00'),
                                     'Powell BART Station, Market
St and Powell St, San Francisco, CA': Timestamp('2019-02-04 00:00:00'))})
```

In [ ]:

```
In [410]: def retro_dictify(frame):
            d = {}
            for row in frame.values:
                here = d
                for elem in row[:-2]:
                    if elem not in here:
                        here[elem] = {}
                    here = here[elem]
                here[row[-2]] = row[-1]
            return d
pprint.pprint(retro_dictify(test_barts_hotspots_filled[['Origin_Name', 'Destination_Name',
'weekday',
'Midday Mean Travel Time',
'Evening Mean Travel Time',
'Early Morning Mean Travel Time']]))
```

```
725.0: {724.0: 617.0},
```

```
735.0: {756.0: 508.0},
```

```
746.0: {743.0: 513.5},
```

```
757.0: {747.0: 423.0},
```

```
760.0: {814.0: 513.5},
```

```
770.0: {718.0: 550.0},
```

```
785.0: {695.0: 497.0},
```

```
797.0: {714.0: 513.5},
```

```
800.0: {932.0: 522.0},
```

```
803.0: {628.0: 513.5},
```



```
In [420]: train_df = barts_hotspots_filled[['Destination Name',
      'Origin Name',
      'Date',
      'weekday',
      'Daily Mean Travel Time']]
train_df['month'] = train_df['Date'].dt.month
```

```
In [422]: train_df.head()
```

```
Out[422]:
```

	Destination Name	Origin Name	Date	weekday	Daily Mean Travel Time	month
0	Embarcadero, San Francisco, CA	The Palace Of Fine Arts, 3601 Lyon St, San Fra...	2019-06-09	6	1588.0	6
1	Fisherman's Wharf, 286-298 Jefferson St, San F...	Embarcadero, San Francisco, CA	2019-04-28	6	639.0	4
2	Fisherman's Wharf, 286-298 Jefferson St, San F...	Embarcadero, San Francisco, CA	2019-05-18	5	730.0	5
3	Embarcadero, San Francisco, CA	Oracle Park, 24 Willie Mays Plaza, San Francis...	2019-04-18	3	512.0	4
4	Embarcadero, San Francisco, CA	Oracle Park, 24 Willie Mays Plaza, San Francis...	2019-04-08	0	505.0	4

```
In [426]:
```

```
In [439]: def train(train_df):
    train_df = train_df.copy()
    X_train, X_test, y_train, y_test = train_test_split(train_df.drop(['Daily Mean Travel Time', 'Date'], axis=1),
                                                        train_df['Daily Mean Travel Time'],
                                                        test_size=0.2,
                                                        random_state=42)

    col_c3 = ['Destination Name', 'Origin Name', 'weekday', 'month']
    c3_transformer = Pipeline(steps=[('onehot', OneHotEncoder())])

    preproc = ColumnTransformer(transformers=[
        ('c3', c3_transformer, col_c3)])

    pl = Pipeline(steps=[('preprocessor', preproc), ('regressor', LinearRegression())])
    pl.fit(X_train, y_train)
    pred = pl.predict(X_test)
    rms = sqrt(mean_squared_error(y_test, pred))
    preds = pl.predict(train_df.drop(['Daily Mean Travel Time', 'Date'], axis=1))
    train_df['pred'] = preds
    return rms, train_df
```

```
In [442]: train(train_df)[1]
```

3240	The Palace Of Fine Arts, 3601 Lyon St, San Francisco, CA 94118	Powell BART Station, Market St and Powell St, San Francisco, CA 94102	2019-03-13	2	1546.0	3	1380.431722
3241	Powell BART Station, Market St and Powell St, San Francisco, CA 94102	The Palace Of Fine Arts, 3601 Lyon St, San Francisco, CA 94118	2019-01-25	4	1536.0	1	1534.961391
3242	Powell BART Station, Market St and Powell St, San Francisco, CA 94102	Fisherman's Wharf, 286-298 Jefferson St, San Francisco, CA 94133	2019-01-22	1	1007.0	1	1052.169465
3243	Fisherman's Wharf, 286-298 Jefferson St, San Francisco, CA 94133	Powell BART Station, Market St and Powell St, San Francisco, CA 94102	2019-01-18	4	933.0	1	920.699240
3244	Powell BART Station, Market St and Powell St, San Francisco, CA 94102	Oracle Park, 24 Willie Mays Plaza, San Francisco, CA 94103	2019-03-25	0	877.0	3	738.279590
3245	Powell BART Station, Market St and Powell St, San Francisco, CA 94102	Oracle Park, 24 Willie Mays Plaza, San Francisco, CA 94103	2019-02-22	4	910.0	2	804.284001
3246	The Palace Of Fine Arts, 3601 Lyon St, San Francisco, CA 94118	Powell BART Station, Market St and Powell St, San Francisco, CA 94102	2019-02-26	1	1462.0	2	1393.802424
3247	Fisherman's Wharf, 286-298 Jefferson St, San Francisco, CA 94133	Powell BART Station, Market St and Powell St, San Francisco, CA 94102	2019-02-13	2	956.0	2	934.156906
3248	Powell BART Station, Market St and Powell St, San Francisco, CA 94102	Oracle Park, 24 Willie Mays Plaza, San Francisco, CA 94103	2019-02-21	3	1029.0	2	817.059705

```
In [448]: barts_hotspots_filled.head()
```

```
Out[448]:
```

	Date	Origin ID	Origin Name	Destination ID	Destination Name	Daily Mean Travel Time	Daily Range - Lower	Daily Range - Upper	AM Mean Travel Time	AM Range - Lower	...	Midday Range - Lower	Midday Range - Upper	Evening Mean Travel Time	Evening Range - Lower	...
0	2019-06-09	3396	The Palace Of Fine Arts, 3601 Lyon St, San Francisco, CA	3603	Embarcadero, San Francisco, CA	1588.0	1054.0	2392.0	NaN	NaN	...	1191.5	1640.5	1149.5	861.0	...
1	2019-04-28	3603	Embarcadero, San Francisco, CA	3394	Fisherman's Wharf, 286-298 Jefferson St, San Francisco, CA	639.0	512.0	796.0	579.0	427.0	...	524.0	817.0	608.0	492.0	...
2	2019-05-18	3603	Embarcadero, San Francisco, CA	3394	Fisherman's Wharf, 286-298 Jefferson St, San Francisco, CA	730.0	520.0	1024.0	567.0	460.0	...	590.0	1061.0	569.0	406.0	...
3	2019-04-18	3792	Oracle Park, 24 Willie Mays Plaza, San Francisco, CA	3603	Embarcadero, San Francisco, CA	512.0	371.0	705.0	648.0	461.0	...	398.0	702.0	442.0	321.0	...
4	2019-04-08	3792	Oracle Park, 24 Willie Mays Plaza, San Francisco, CA	3603	Embarcadero, San Francisco, CA	505.0	364.0	701.0	621.0	489.0	...	339.0	542.0	502.0	354.0	...

5 rows × 25 columns

```
In [449]: barts_hotspots_filled['Origin Name'].unique()
```

```
Out[449]: array(['The Palace Of Fine Arts, 3601 Lyon St, San Francisco, CA',
                  'Embarcadero, San Francisco, CA',
                  'Oracle Park, 24 Willie Mays Plaza, San Francisco, CA',
                  'Fisherman's Wharf, 286-298 Jefferson St, San Francisco, CA',
                  '2nd Street and Stevenson Street, San Francisco, CA',
                  'Powell BART Station, Market St and Powell St, San Francisco, CA'],
              dtype=object)
```

```
In [457]: temp = elvis[(barts_hotspots_filled['Origin Name'] == "2nd Street and Stevenson Street, San Francisco, CA") && (temp > 0)]
```

				St, San Fra...			Time	Lower	Upper		Upper	Time	Lower	
686	686	2019-06-21	3692	The Palace Of Fine Arts, 3601 Lyon St, San Fra...	3396	The Palace Of Fine Arts, 3601 Lyon St, San Fra...	1384.0	1017.0	1885.0	1285.000000	...	1602.0	1146.000000	975.000000
692	692	2019-05-02	3692	The Palace Of Fine Arts, 3601 Lyon St, San Fra...	3396	The Palace Of Fine Arts, 3601 Lyon St, San Fra...	1241.0	901.0	1711.0	1321.000000	...	1833.0	1178.000000	893.000000

```
In [459]: temp[temp['weekday']==1]
```

id	Name	Travel Time	Travel Time Lower	Travel Time Upper	Travel Time	...	Travel Time Upper	Travel Time	Travel Time Lower	Travel Time Upper	Mean Travel Time	Range Lower	Range Upper	weekday	pred_daily
3396	Arts, 3601 Lyon St, San Fra...	1265.0	974.0	1644.0	1143.408226	...	1639.0	1150.0	981.0	1388.0	733.432934	560.000000	952.000000	1	1223.2
3396	The Palace Of Fine Arts, 3601 Lyon St, San Fra...	1265.0	978.0	1635.0	1143.408226	...	1689.5	1150.0	916.0	1444.0	733.432934	560.000000	952.000000	1	1213.8
3396	The Palace Of Fine Arts, 3601 Lyon St,	1309.0	990.0	1730.0	1256.907919	...	1767.0	1151.0	981.0	1414.0	740.000000	576.000000	978.000000	1	1223.2

```
In [474]: elvis = pd.read_csv("data/out.csv")
          elvis
```

```
it[474]:
```

	Unnamed: 0	Date	Origin ID	Origin Name	Destination ID	Destination Name	Daily Mean Travel Time	Daily Range - Lower	Daily Range - Upper	AM Mean Travel Time	...	Midday Range - Upper	Evening Mean Travel Time	Evening Range - Lower
0	0	2019-06-09	3396	The Palace Of Fine Arts, 3601 Lyon St, San Fra...	3603	Embarcadero, San Francisco, CA	1588.0	1054.0	2392.0	1580.000000	...	1640.5	1149.500000	861.0
1	1	2019-04-28	3603	Embarcadero, San Francisco, CA	3394	Fisherman's Wharf, 286-298 Jefferson St, San F...	639.0	512.0	796.0	579.000000	...	817.0	608.000000	492.0
2	2	2019-05-18	3603	Embarcadero, San Francisco, CA	3394	Fisherman's Wharf, 286-298 Jefferson St, San F...	730.0	520.0	1024.0	567.000000	...	1061.0	569.000000	406.0
3	3	2019-04-18	3792	Oracle Park, 24 Willie Mays Plaza	3603	Embarcadero, San	512.0	371.0	705.0	648.000000	...	702.0	442.000000	321.0

```
In [ ]:
```