```
In [2]:  import pandas as pd
         from postal.expand import expand_address
         from postal.parser import parse_address
```

```
In [ ]:  mordor_zagats_columns = ['id', 'name', 'address', 'city', 'phone', 'cat
```

```
In [36]: yellow_yelp = pd.read_csv(
             'yellow_yelp_all.csv',
             sep=',',
             quotechar='"',
             )
```

```
In [38]: def expand_address_value(address_field):
             expanded_address = expand_address(str(address_field), languages = [

             if len(expanded_address) == 1:
                 return expanded_address[0]
             else:
                 return address_field
```

```
In [39]: yellow_yelp['exp_addr'] = yellow_yelp['address'].apply(lambda x: expand
```

```
In [6]:  def normalize_addr(entry):

             addr_to_parse =  entry['exp_addr']  #+ ',' + ' ' * 1 + entry['city'

             x = parse_address(
                 addr_to_parse,
                 language='en',
                 country='us')

             for val_combo in x:
                 column_val = val_combo[0]
                 column_name = val_combo[1]
                 full_col_name = 'pypost_' + column_name
                 entry[full_col_name] = column_val

             return entry
```

In [40]: `normalize_addr(yellow_yelp.iloc[0])`

/home/boris/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:14: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy (http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy)

Out[40]:
```
Unnamed: 0                             0
id                                     1
name                   Tao Tao Restaurant
address                  175 S Murphy Ave
city                           Sunnyvale
postalcode                         94086
phone                     (408) 736-3731
source                      yellow_pages
exp_addr                 175 S Murphy Ave
pypost_house_number                  175
pypost_road                 s murphy ave
Name: 0, dtype: object
```

In [41]: `df2 = yellow_yelp.apply(lambda x: normalize_addr(x), axis=1)`

```python
In [28]: def add_new_pp_info(entry):

             entry['street'] = None
             entry['house_number'] = None
             entry['house'] = None

             entry['address'] = entry['exp_addr']

             if entry['pypost_road'] is not None:
                 entry['house_number'] = entry['pypost_house_number']
                 entry['house'] = entry['pypost_house']
                 entry['street'] = entry['pypost_road']


             fields_of_interest = [
                 'id',
                 'source',
                 'name',
                 'category',
                 'phone',
                 'city',
                 'postalcode',
                 'address',
                 'street',
                 'house_number',
                 'house',

             ]


             return entry[fields_of_interest]
```

```python
In [43]: final_df = df2.apply(lambda x: add_new_pp_info(x), axis=1)
```

```
In [44]: final_df.head()
```

Out[44]:

	id	source	name	category	phone	city	postalcode	address	street	hous
0	1	yellow_pages	Tao Tao Restaurant	NaN	(408) 736-3731	Sunnyvale	94086.0	175 S Murphy Ave	s murphy ave	
1	2	yellow_pages	Dinner at Buca di Beppo	NaN	NaN	NaN	NaN	nan	NaN	
2	3	yellow_pages	Sneha South & North Indian Restaurant	NaN	(408) 481-0700	Sunnyvale	94085.0	1214 Apollo Way # 404-B	apollo way	
3	4	yellow_pages	The Armenian Gourmet	NaN	(408) 732-3910	Sunnyvale	94085.0	929 E Duane Ave	e duane ave	
4	5	yellow_pages	Round Table Pizza	NaN	(408) 733-1365	Sunnyvale	94085.0	415 N Mary Ave	n mary ave	

```
In [46]: final_df.to_csv('yellow_yelp_all_pypostal.csv', sep= ',', quotechar = '
```