```
In [10]:  import pandas as pd
          from postal.expand import expand_address

          from postal.expand import *
          expand_address('Eins Herbert-von-Karajan-Str. 10785 Berlin', address_com
```

Out[10]: ['1 herbert-von-karajanstrasse 10785 berlin',
          '1 herbert-von-karajan strasse 10785 berlin',
          '1 herbert von karajanstrasse 10785 berlin',
          '1 herbert von karajan strasse 10785 berlin',
          '1 herbertvonkarajanstrasse 10785 berlin',
          '1 herbertvonkarajan strasse 10785 berlin']

```
In [6]:  from postal.parser import parse_address
```

```
In [9]:  parse_address('1 herbert-von-karajanstrasse 10785 berlin')
```

Out[9]: [('1', 'house_number'),
          ('herbert-von-karajanstrasse', 'road'),
          ('10785', 'postcode'),
          ('berlin', 'city')]

```
In [12]:  relevant_columns = ['Id', 'name', 'address', 'city', 'phone', 'category
```

```
In [15]:  fodors_zagats = pd.read_csv(
              'fodors_zagats_all.csv',
              sep=',',
              quotechar='"',
              )relevant_columns
```

```
In [16]:  fodors_zagats.head()
```

Out[16]:

	Unnamed: 0	Id	name	address	city	phone	category	source
0	0	534	arnie morton\s of chicago'	435 s. la cienega blv.	los angeles	310/246-1501	american	fodors
1	1	535	art\s delicatessen'	12224 ventura blvd.	studio city	818/762-1221	american	fodors
2	2	536	hotel bel-air	701 stone canyon rd.	bel air	310/472-1211	californian	fodors
3	3	537	cafe bizou	14016 ventura blvd.	sherman oaks	818/788-3536	french	fodors
4	4	538	campanile	624 s. la brea ave.	los angeles	213/938-1447	american	fodors

```python
In [118]: 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
```

```python
In [119]: fodors_zagats['exp_addr'] = fodors_zagats['address'].apply(lambda x: ex|
```

```python
In [120]: fodors_zagats.head(3)
```

Out[120]:

	Unnamed: 0	Id	name	address	city	phone	category	source	exp_addr
0	0	534	arnie morton\s of chicago'	435 s. la cienega blv.	los angeles	310/246-1501	american	fodors	435 s. la cienega blv.
1	1	535	art\s delicatessen'	12224 ventura blvd.	studio city	818/762-1221	american	fodors	12224 ventura boulevard
2	2	536	hotel bel-air	701 stone canyon rd.	bel air	310/472-1211	californian	fodors	701 stone canyon road

```python
In [168]: 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 [126]: normalize_addr(fodors_zagats.iloc[0])
```

/home/boris/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.p
y: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.pyd
ata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy)

```
Out[126]: Unnamed: 0                                    0
          Id                                          534
          name                    arnie morton\s of chicago'
          address                    435 s. la cienega blv.
          city                              los angeles
          phone                             310/246-1501
          category                              american
          source                                  fodors
          exp_addr                   435 s. la cienega blv.
          pypost_house_number                         435
          pypost_road                    s. la cienega blv.
          Name: 0, dtype: object
```

```
In [127]: df2 = fodors_zagats.apply(lambda x: normalize_addr(x), axis=1)
```

```
In [193]: 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',
                  'address',
                  'street',
                  'house_number',
                  'house',

              ]


              return entry[fields_of_interest]
```

```
In [194]: final_df = df2.apply(lambda x: add_new_pp_info(x), axis=1)
```

```
In [196]: final_df.head()
```

Out[196]:

	Id	source	name	category	phone	city	address	street	house_number
0	534	fodors	arnie morton\s of chicago'	american	310/246-1501	los angeles	435 s. la cienega blv.	s. la cienega blv.	435
1	535	fodors	art\s delicatessen'	american	818/762-1221	studio city	12224 ventura boulevard	ventura boulevard	12224
2	536	fodors	hotel bel-air	californian	310/472-1211	bel air	701 stone canyon road	stone canyon road	701
3	537	fodors	cafe bizou	french	818/788-3536	sherman oaks	14016 ventura boulevard	ventura boulevard	14016
4	538	fodors	campanile	american	213/938-1447	los angeles	624 s. la brea ave.	s. la brea ave.	624

```
In [197]: final_df.to_csv('fodors_zagats_pypostal.csv', sep= ',', quotechar = '"'
```