In [1]:
```python
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```python
house_hold_df = pd.read_excel("C:/Users/Eng Farhad/Desktop/Farhad Achikzai Acted Dataset/dataset/input/V03_translated_-_all_versions.xlsx",
                              sheet_name='data')
```

In [3]:
```python
house_hold_df.head(3)
```

Out[3]:

	province	district	instance_name	repeat_in_c_2	Household	pop_group_condition	hh_move_why	hh_demographics_note	hh_resp_gender	hh_resp_hoh	...	hh_income_fluctuation	hh_debt
0	AF09	AF0902	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	5	h_1	displaced	anticipated_change	NaN	male	yes	...	no_higher	yes
1	AF33	AF3301	bb1640ab-db0f-4fa4-b261-5c93721026d8	6	h_2	host	NaN	NaN	male	yes	...	yes	yes
2	AF09	AF0902	ea95be91-1d5f-4524-8346-d6090fee185d	2	h_3	displaced	poverty_unemployment	NaN	male	yes	...	yes	yes

3 rows × 33 columns

In [4]:
```python
house_hold_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13 entries, 0 to 12
Data columns (total 33 columns):
 #   Column                      Non-Null Count  Dtype  
---  ------                      --------------  -----  
 0   province                    13 non-null     object 
 1   district                    13 non-null     object 
 2   instance_name               13 non-null     object 
 3   repeat_in_c_2               13 non-null     int64  
 4   Household                   13 non-null     object 
 5   pop_group_condition         13 non-null     object 
 6   hh_move_why                 8 non-null      object 
 7   hh_demographics_note        0 non-null      float64
 8   hh_resp_gender              13 non-null     object 
 9   hh_resp_hoh                 13 non-null     object 
 10  hh_hoh_gender_calc_1        13 non-null     object 
 11  hh_latrine                  13 non-null     object 
 12  energy_lighting             13 non-null     object 
 13  energy_com_fac              13 non-null     object 
 14  hh_breadwinner_male         13 non-null     int64  
 15  change_employment           7 non-null      object 
 16  hh_displaced_income         11 non-null     object 
 17  hh_displaced_income_other   0 non-null      float64
 18  hh_business_availability    13 non-null     object 
 19  hh_business_sector          6 non-null      object 
 20  hh_business_sector_other    0 non-null      float64
 21  hh_business_networks        6 non-null      object 
 22  hh_income                   12 non-null     float64
 23  hh_income_fluctuation       13 non-null     object 
 24  hh_debt                     13 non-null     object 
 25  debt_exp                    12 non-null     object 
 26  other_exp                   13 non-null     int64  
 27  total_exp                   13 non-null     object 
 28  pov_line_calculation        13 non-null     object 
 29  hh_land_access              12 non-null     object 
 30  hh_tazkera                  13 non-null     int64  
 31  _uuid                       13 non-null     object 
 32  _index                      13 non-null     int64  
dtypes: float64(4), int64(5), object(24)
memory usage: 3.5+ KB
```

In [5]:
```python
houshold_sheet2 = pd.read_excel("C:/Users/Eng Farhad/Desktop/Farhad Achikzai Acted Dataset/dataset/input/V03_translated_-_all_versions.xlsx",
                              sheet_name='c_2')
```

In [6]:
```python
houshold_sheet2.head(2)
```

Out[6]:

	parent_instance_name	person_id	household_member_sex	_parent_index	_submission__uuid
0	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	fcf259cb-8045-4454-83c4-48d2eb9d8b13	male	1	74f6150e-7cd2-4bdb-9bf2-3de7f4c82b30
1	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	a51ee452-b0ba-4a7f-8a7e-12b4f52e0c38	male	1	74f6150e-7cd2-4bdb-9bf2-3de7f4c82b30

In [7]:
```python
#merging data from multiple table for analysis
house_hold_join = pd.merge(house_hold_df, houshold_sheet2,
                how='right', left_on='instance_name', right_on='parent_instance_name')
```

In [8]:
```python
house_hold_join.head(3)
```

Out[8]:

	province	district	instance_name	repeat_in_c_2	Household	pop_group_condition	hh_move_why	hh_demographics_note	hh_resp_gender	hh_resp_hoh	...	pov_line_calculation	hh_land_access
0	AF09	AF0902	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	5	h_1	displaced	anticipated_change	NaN	male	yes	...	below_pov	no
1	AF09	AF0902	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	5	h_1	displaced	anticipated_change	NaN	male	yes	...	below_pov	no
2	AF09	AF0902	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	5	h_1	displaced	anticipated_change	NaN	male	yes	...	below_pov	no

3 rows × 38 columns

In [9]:
```python
# Percentage % of income injected in the district
round(house_hold_join.groupby('district')['hh_income'].sum()*100/house_hold_join['hh_income'].sum(), 2)
```

Out[9]:

```
district
AF0109     0.03
AF0505    99.87
AF0702     0.01
AF0902     0.01
AF0905     0.02
AF1001     0.00
AF1005     0.01
AF1501     0.02
AF3001     0.01
AF3202     0.00
AF3301     0.00
Name: hh_income, dtype: float64
```

In [10]:
```python
# Percentage % of income received by community compare to residency (host or displaced) community
round(house_hold_join[['pop_group_condition', 'hh_income']].
      groupby('pop_group_condition').sum()*100/house_hold_join['hh_income'].sum(), 2)
```

Out[10]:

pop_group_condition	hh_income
displaced	99.99
host	0.01

In [11]:
```python
#sum of amount dedicated to province
house_hold_join.groupby('province')['hh_income'].sum()
```

Out[11]:

```
province
AF01        67200.0
AF05    200000000.0
AF07        20000.0
AF09        62596.0
AF10        33000.0
AF15        45000.0
AF30        25000.0
AF32            0.0
AF33         3996.0
Name: hh_income, dtype: float64
```

In [12]:
```python
#percentage% of amount dedicated to the provinces
round(house_hold_join[['province', 'hh_income']].groupby('province').sum()*100/house_hold_join['hh_income'].sum(), 3)
```

Out[12]:

province	hh_income
AF01	0.034
AF05	99.872
AF07	0.010
AF09	0.031
AF10	0.016
AF15	0.022
AF30	0.012
AF32	0.000
AF33	0.002

In [13]:
```python
house_hold_join.head(3)
```

Out[13]:

	province	district	instance_name	repeat_in_c_2	Household	pop_group_condition	hh_move_why	hh_demographics_note	hh_resp_gender	hh_resp_hoh	...	pov_line_calculation	hh_land_access
0	AF09	AF0902	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	5	h_1	displaced	anticipated_change	NaN	male	yes	...	below_pov	no
1	AF09	AF0902	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	5	h_1	displaced	anticipated_change	NaN	male	yes	...	below_pov	no
2	AF09	AF0902	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	5	h_1	displaced	anticipated_change	NaN	male	yes	...	below_pov	no

3 rows × 38 columns

In [14]:
```python
#Amount was distributed by Gender
house_hold_join.groupby('household_member_sex')['hh_income'].sum()
```

Out[14]:

```
household_member_sex
female    100087266.0
male      100169526.0
Name: hh_income, dtype: float64
```

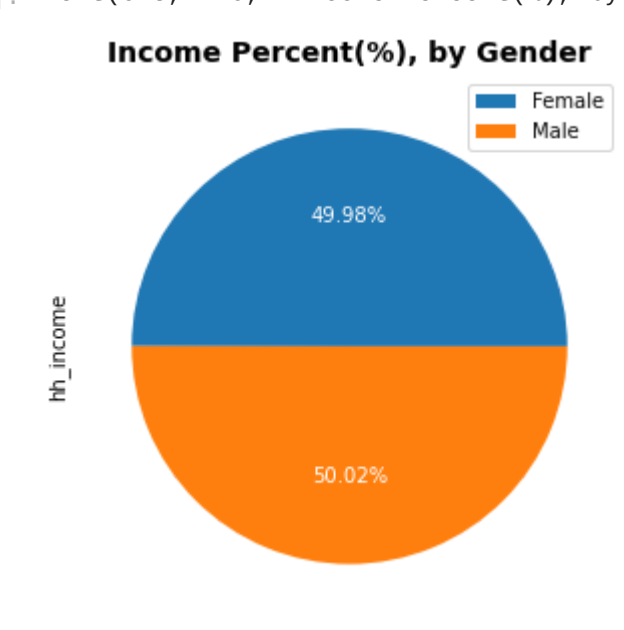
In [15]:
```python
# percentage% of amount distributed by gender

round(house_hold_join[['household_member_sex', 'hh_income']].
      groupby('household_member_sex').sum()*100/house_hold_join['hh_income'].sum(), 2)
```

Out[15]:

household_member_sex	hh_income
female	49.98
male	50.02

In [16]:
```python
#income Percentage% by sex
legend = ['Female', 'Male']
house_hold_join.groupby('household_member_sex')['hh_income'].sum().plot(kind='pie',
                                                                          autopct='%.2f%%',
                                                                          textprops={'color':'w'}, figsize=(6,5))
plt.legend(legend, loc='upper right')
plt.title('Income Percent(%), by Gender', fontsize=14, fontweight='bold')
```

Out[16]: Text(0.5, 1.0, 'Income Percent(%), by Gender')



In [17]:
```python
# percentage% of poverity among the community
round(house_hold_join[['pov_line_calculation', 'hh_income']].
      groupby('pov_line_calculation').count() * 100 / house_hold_join['pov_line_calculation'].count(), 2)
```

Out[17]:

pov_line_calculation	hh_income
above_pov	19.64
below_pov	75.00

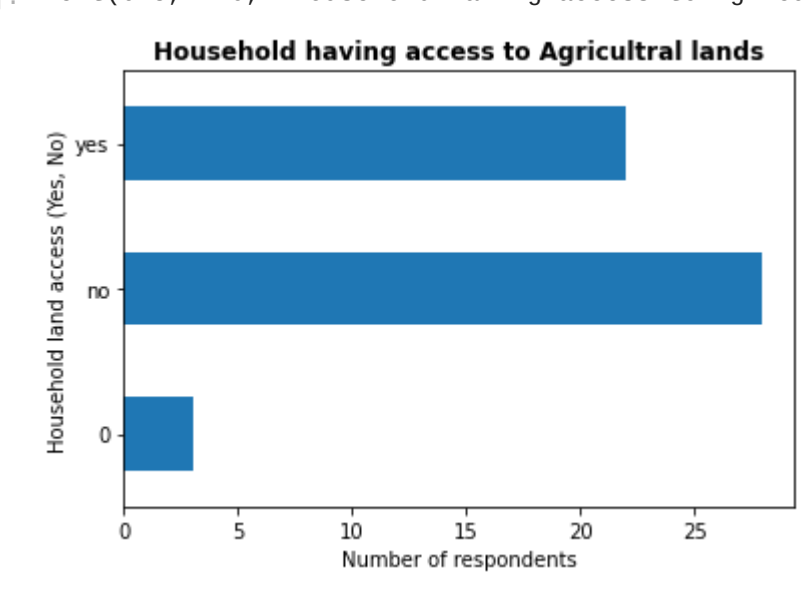
In [18]:
```python
# percentage % of agricultural land among the community
round(house_hold_join.groupby('hh_land_access')['hh_land_access'].
      count() * 100 / house_hold_join['hh_land_access'].count(),2)
```

Out[18]:

```
hh_land_access
0       5.66
no     52.83
yes    41.51
Name: hh_land_access, dtype: float64
```

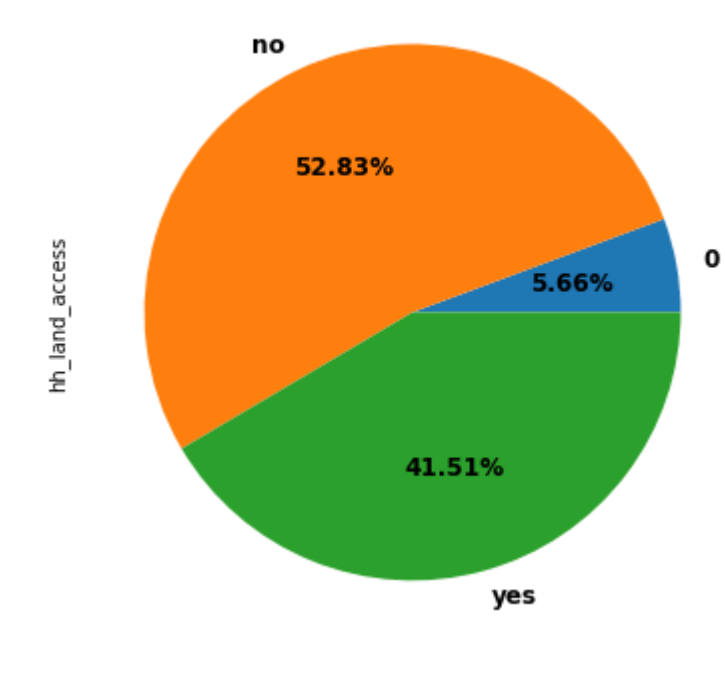
In [19]:
```python
#graph presentation of land accessing in the community
house_hold_join.groupby('hh_land_access')['hh_land_access'].count().plot(kind='barh')
plt.ylabel('Household land access (Yes, No)')
plt.xlabel('Number of respondents')
plt.title('Household having access to Agricultral lands', fontsize=12, fontweight='bold')
```

Out[19]: Text(0.5, 1.0, 'Household having access to Agricultral lands')



In [20]:
```python
#household accessing agricultural land per%
house_hold_join.groupby('hh_land_access')['hh_land_access'].count().plot(kind='pie',
                                                                          autopct='%.2f%%',
                                                                          figsize=(6,8),
                                                                          textprops={'fontsize':12, 'fontweight':'bold'})
```

Out[20]: <AxesSubplot:ylabel='hh_land_access'>



In [21]:
```python
house_hold_join.head(3)
```

Out[21]:

	province	district	instance_name	repeat_in_c_2	Household	pop_group_condition	hh_move_why	hh_demographics_note	hh_resp_gender	hh_resp_hoh	...	pov_line_calculation	hh_land_access
0	AF09	AF0902	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	5	h_1	displaced	anticipated_change	NaN	male	yes	...	below_pov	no
1	AF09	AF0902	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	5	h_1	displaced	anticipated_change	NaN	male	yes	...	below_pov	no
2	AF09	AF0902	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	5	h_1	displaced	anticipated_change	NaN	male	yes	...	below_pov	no

3 rows × 38 columns

In [22]:
```python
house_hold_join['hh_land_access'].unique()
```

Out[22]: array(['no', 'yes', 0, nan], dtype=object)

In [23]:
```python
#grouping of unwilling to response and null value into (unkonw)
house_hold_join['hh_land_access'] = house_hold_join['hh_land_access'].replace(0,'unknow')
```

In [24]:
```python
house_hold_join['hh_move_why'].unique()
```

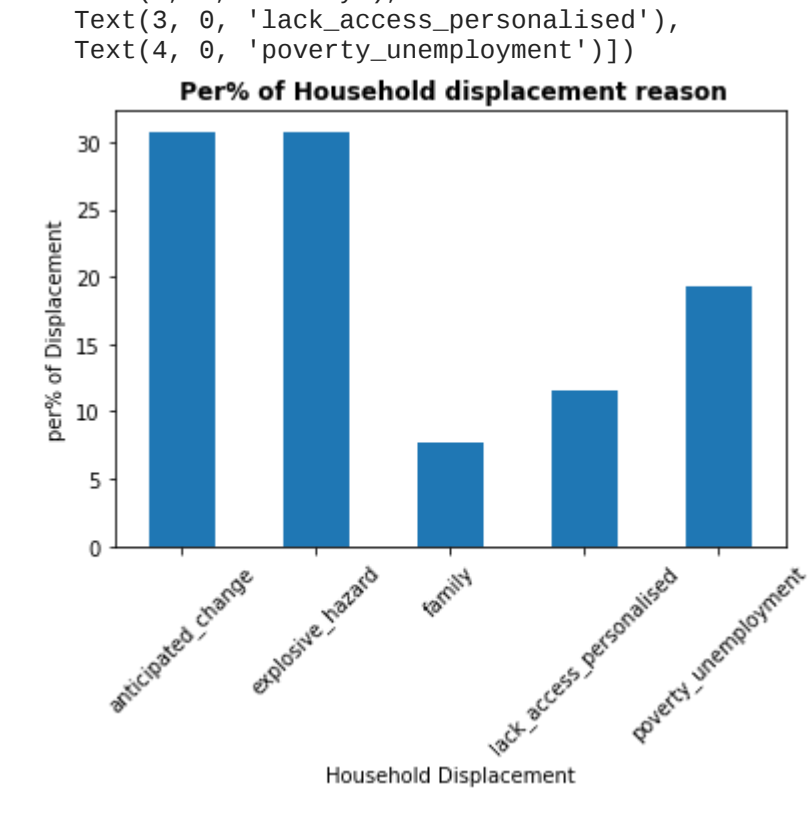
Out[24]:

```
array(['anticipated_change', nan, 'poverty_unemployment',
       'explosive_hazard', 'family', 'lack_access_personalised'],
      dtype=object)
```

In [25]:
```python
# percentage % of household displacement from their homland
round(house_hold_join.groupby('hh_move_why')['hh_move_why'].
        count()*100/house_hold_join['hh_move_why'].count(),2).plot(kind='bar')
plt.xlabel('Household Displacement')
plt.ylabel('per% of Displacement')
plt.title(' Per% of Household displacement reason', fontweight='bold')
plt.xticks(rotation=45)
```

Out[25]:

```
(array([0, 1, 2, 3, 4]),
 [Text(0, 0, 'anticipated_change'),
  Text(1, 0, 'explosive_hazard'),
  Text(2, 0, 'family'),
  Text(3, 0, 'lack_access_personalised'),
  Text(4, 0, 'poverty_unemployment')])
```



In [26]:
```python
corr = house_hold_join.corr()
```

In [27]:
```python
corr
```

Out[27]:

	repeat_in_c_2	hh_demographics_note	hh_breadwinner_male	hh_displaced_income_other	hh_business_sector_other	hh_income	other_exp	hh_tazkera	_index	_parent_in
repeat_in_c_2	1.000000	NaN	-0.516278	NaN	NaN	-0.244165	-0.288869	0.515289	-0.118537	-0.118
hh_demographics_note	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
hh_breadwinner_male	-0.516278	NaN	1.000000	NaN	NaN	-0.130167	-0.186222	0.050233	0.270003	0.270
hh_displaced_income_other	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
hh_business_sector_other	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
hh_income	-0.244165	NaN	-0.130167	NaN	NaN	1.000000	-0.022826	-0.024057	-0.032885	-0.032
other_exp	-0.288869	NaN	-0.186222	NaN	NaN	-0.022826	1.000000	-0.302023	0.383956	0.383
hh_tazkera	0.515289	NaN	0.050233	NaN	NaN	-0.024057	-0.302023	1.000000	0.427170	0.427
_index	-0.118537	NaN	0.270003	NaN	NaN	-0.032885	0.383956	0.427170	1.000000	1.000
_parent_index	-0.118537	NaN	0.270003	NaN	NaN	-0.032885	0.383956	0.427170	1.000000	1.000

In [28]:
```python
house_hold_join.head(3)
```

Out[28]:

	province	district	instance_name	repeat_in_c_2	Household	pop_group_condition	hh_move_why	hh_demographics_note	hh_resp_gender	hh_resp_hoh	...	pov_line_calculation	hh_land_access
0	AF09	AF0902	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	5	h_1	displaced	anticipated_change	NaN	male	yes	...	below_pov	no
1	AF09	AF0902	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	5	h_1	displaced	anticipated_change	NaN	male	yes	...	below_pov	no
2	AF09	AF0902	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	5	h_1	displaced	anticipated_change	NaN	male	yes	...	below_pov	no

3 rows × 38 columns

In [34]:
```python
house_hold_join['hh_move_why'].unique()
```

Out[34]:

```
array(['anticipated_change', nan, 'poverty_unemployment',
       'explosive_hazard', 'family', 'lack_access_personalised'],
      dtype=object)
```

In [35]:
```python
group_A = (['anticipated_change', 'nan', 'family'])
group_B = (['explosive_hazard', 'poverty_unemployment', 'lack_access_personalised'])
```

In [36]:
```python
group_A
```

Out[36]: ['anticipated_change', 'nan', 'family']

In [ ]:
```python
 house_hold_join['family_move'] =  house_hold_join['family_move'].replace('A', 'Family problem')
```

In [ ]:
```python
house_hold_join['family_move'] = house_hold_join['family_move'].replace('B', 'Economical problem')
```

In [ ]:
```python
round(house_hold_join.groupby('family_move')['family_move'].
      count()*100/house_hold_join['family_move'].count(), 2).plot(kind='bar', figsize=(6,6))
plt.xlabel('Family Displacement')
plt.ylabel('%percentage of Displacement')
plt.title('Family Displacement % for economical or family problems', fontweight='bold')
plt.xticks(rotation=15)
```

In [38]:
```python
house_hold_join.head(2)
```

Out[38]:

	province	district	instance_name	repeat_in_c_2	Household	pop_group_condition	hh_move_why	hh_demographics_note	hh_resp_gender	hh_resp_hoh	...	pov_line_calculation	hh_land_access
0	AF09	AF0902	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	5	h_1	displaced	anticipated_change	NaN	male	yes	...	below_pov	no
1	AF09	AF0902	fd0c1fdb-7793-4f3a-bd58-439c39b23a4f	5	h_1	displaced	anticipated_change	NaN	male	yes	...	below_pov	no

2 rows × 38 columns

In [39]:
```python
house_hold_join.groupby('province')[['hh_income', 'total_exp']].sum()
```

Out[39]:

province	hh_income
AF01	67200.0
AF05	200000000.0
AF07	20000.0
AF09	62596.0
AF10	33000.0
AF15	45000.0
AF30	25000.0
AF32	0.0
AF33	3996.0

In [40]:
```python
house_hold_join.groupby("district")[["hh_income", "total_exp"]].sum()
```

Out[40]:

district	hh_income
AF0109	67200.0
AF0505	200000000.0
AF0702	20000.0
AF0902	26596.0
AF0905	36000.0
AF1001	3000.0
AF1005	30000.0
AF1501	45000.0
AF3001	25000.0
AF3202	0.0
AF3301	3996.0

In [41]:
```python
house_hold_join.groupby(["hh_income", "total_exp"]).sum()
```

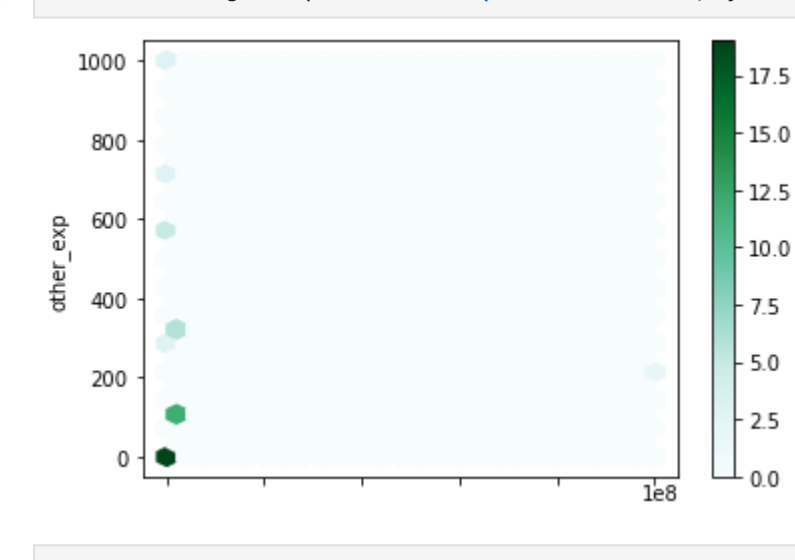
Out[41]:

hh_income	total_exp	repeat_in_c_2	hh_demographics_note	hh_breadwinner_male	hh_displaced_income_other	hh_business_sector_other	other_exp	hh_tazkera	_index	_parent_index
666.0	10457	36	0.0	6	0.0	0.0	1920	-18	12	12
1000.0	100	9	0.0	6	0.0	0.0	0	6	27	27
	5149	32	0.0	4	0.0	0.0	0	8	16	16
1298.0	23902	4	0.0	4	0.0	0.0	200	2	6	6
4000.0	10642	25	0.0	10	0.0	0.0	495	10	5	5
	13266	45	0.0	5	0.0	0.0	2830	25	60	60
5000.0	4750	25	0.0	5	0.0	0.0	500	10	50	50
5600.0	14600	144	0.0	12	0.0	0.0	0	72	84	84
10000.0	21100	9	0.0	3	0.0	0.0	2100	-9	24	24
12000.0	12800	9	0.0	6	0.0	0.0	900	6	15	15
15000.0	14500	3	0.0	6	0.0	0.0	3000	9	39	39
100000000.0	exp	4	0.0	2	0.0	0.0	400	4	12	12

In [ ]:
```python
#house_hold_join.plot.hexbin(x=house_hold_join['hh_income'], y=house_hold_join['total_exp'])
```

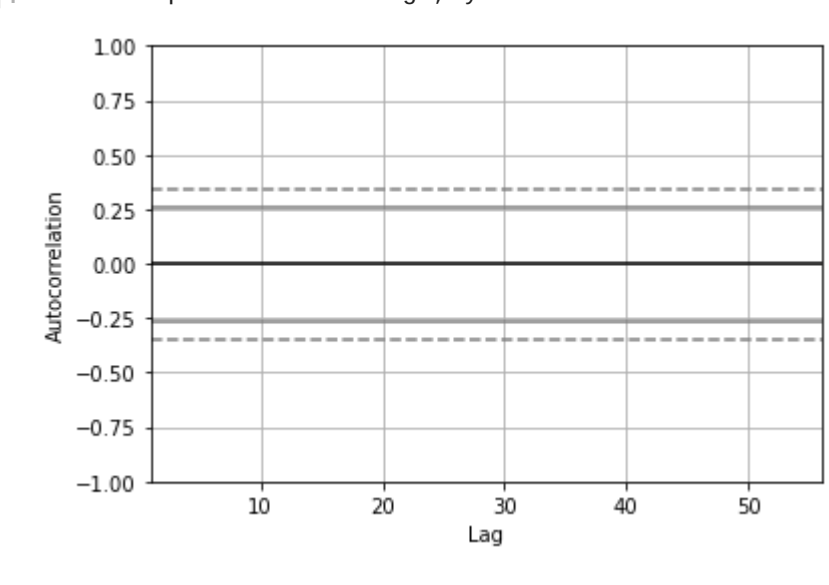
In [42]:
```python
house_hold_join['total_exp'] = house_hold_join['total_exp'].replace('exp', 0)
```

In [43]:
```python
house_hold_join.plot.hexbin(x="hh_income", y="other_exp", gridsize=25);
```



In [44]:
```python
from pandas.plotting import autocorrelation_plot
plt.figure()
autocorrelation_plot(house_hold_join['hh_income'])
```

Out[44]: <AxesSubplot:xlabel='Lag', ylabel='Autocorrelation'>



In [ ]:

In [ ]: