

## Voting booth data analysis

TODO - Talk about EDA

TODO - Talk about the dataset(source,what it contains,how it will be useful)

- Company
- Voters information

## Getting the dataset

Importing pandas to read dataset

	RespondentName	RespondentNumber	Gender	Caste	Age_Group	Education	Occupation	DistrictId	DistrictName	ConstituencyId	...	What are the major sampraday or Gurus followed locally? (Swami Narayan, AsaramBapu etc)	Which tribal related movement is still an issue for upcoming elections in this village?	Which party does the sarpanch/councillor belong to?	Any other factors which determine the winner in upcoming assembly elections in this village?	1_Indicate the support for party among top 3 castes.	2_Indicate the support for party among top 3 castes.	3_Indicate the support for party among top 3 castes.	Is the patidar agitation for reservation still an issue for upcoming elections in this village ?	What are the failures of the sitting MLA	What are the strengths of the sitting MLA
0	NAROTMJI HARJIJI THAKOR	9723281796	Male	Other Service OBCs	26 - 35	Below Primary	Agricultural labourers	2	Banaskantha	7	...	No particular sect	NaN	BJP	NaN	INC Clear	INC Clear	BJP Tight	NaN	Failed to develop constituency	Very accessible
1	KRISHNALAL GOVINDBHAI CHAUDHARI	9687821743	Male	Patel	60 - Above	Intermediate / No Degree	Rentire	24	Navsari	177	...	Parmatama Samaj	NaN	INC	NaN	INC Tight	INC Tight	INC Clear	NaN	Corruption	Very accessible
2	NARENDRASINH NATUBHA VAGHELA	9998873074	Male	Rajput	36 - 45	Graduate or Equivalent	Class-III Employee (Clerical)	9	Rajkot	75	...	Ramapir Dharm	NaN	BJP	NaN	BJP Clear	BJP Clear	INC Tight	NaN	Others	Community
3	SHANTIDAS ISHWARDAS DANIDHARIYA	9879076618	Male	Chaudhary	46 - 60	Professional Degrees and Higher Research	Modern Artist	8	Surendranagar	63	...	Meldimata	NaN	BJP	NaN	BJP Clear	BJP Clear	BJP Tight	NaN	Failed to develop constituency	Local development works
4	DARJIBHAI DHARMSIBHAI PATEL	9913343880	Male	Patel	60 - Above	Non-Literate	Other Agricultural Workers	7	Ahmedabad	40	...	Muslims	NaN	INC	NaN	INC Tight	INC Clear	BJP Clear	NaN	Failed to develop constituency	Community

5 rows × 44 columns

## Data Preparation and cleaning

- Looking at some information about the data and the columns of data set
- Create a new dataset with important columns which are necessary for analysis:
  - Demographics of the survey respondents
  - Dltribution of residents,employment and preserences
- Fix any missing or incorrect values

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1927 entries, 0 to 1926
Data columns (total 44 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   RespondentName                           1927 non-null   object
1   RespondentNumber                         1927 non-null   int64
2   Gender                                   1927 non-null   object
3   Caste                                    1927 non-null   object
4   Age_Group                               1927 non-null   object
5   Education                               1927 non-null   object
6   Occupation                               1927 non-null   object
7   DistrictId                              1927 non-null   int64
8   DistrictName                             1927 non-null   object
9   ConstituencyId                           1927 non-null   int64
10  ConstituencyName                         1927 non-null   object
11  VillageId                               1927 non-null   int64
12  VillageName                             1927 non-null   object
13  BoothId                                 1927 non-null   int64
14  SurveyTime                              1927 non-null   object
15  List out the important  BJP leaders in this village. Also mention if there is friction.  1915 non-null   object
16  List out the important INC leaders in this village. Also mention if there is friction.  1853 non-null   object
17  List out the important independent leaders in this village (AAP, NCP, JDU etc)         1364 non-null   object
18  List out any neutral/unbiased social workers who hold sway                         1532 non-null   object
19  List out the probable MLA candidates for BJP.                                     1901 non-null   object
20  List out the probable MLA candidates for INC                                       1826 non-null   object
21  List out the probable INDEPENDENT MLA candidates.                               1371 non-null   object
22  Can sitting MLA win?                                                              1927 non-null   object
23  1_What are the major public demands of this village?                             59 non-null    object
24  2_What are the major public demands of this village?                             59 non-null    object
25  3_What are the major public demands of this village?                             59 non-null    object
26  Which party has an edge in this village?                                          1927 non-null   object
27  1_List out the important caste groups in descending order population in this village  1927 non-null   object
28  2_List out the important caste groups in descending order population in this village  1927 non-null   object
29  3_List out the important caste groups in descending order population in this village  1927 non-null   object
30  Has the BJP MLA/MLA Candidate visited the village in recent past? (Last 3 months).  1927 non-null   object
```

```
31 Has the INC MLA/MLA Candidate visited the village in recent past? (Last 3 months).      1927 non-null    object
32 What are the major concerns of farmers in this village?                             59 non-null     object
33 What are the major concerns of youth in this village?                             1927 non-null   object
34 What are the major sampraday or Gurus followed locally? (Swami Narayan, AsaramBapu etc) 1927 non-null   object
35 Which tribal related movement is still an issue for upcoming elections in this village? 0 non-null      float64
36 Which party does the sarpanch/councillor belong to?                               1927 non-null   object
37 Any other factors which determine the winner in upcoming assembly elections in this village? 59 non-null     object
38 1_Indicate the support for party among top 3 castes.                             1927 non-null   object
39 2_Indicate the support for party among top 3 castes.                             1927 non-null   object
40 3_Indicate the support for party among top 3 castes.                             1927 non-null   object
41 Is the patidar agitation for reservation still an issue for upcoming elections in this village ? 49 non-null     object
42 What are the failures of the sitting MLA                                           1868 non-null   object
43 What are the strengths of the sitting MLA                                           1868 non-null   object
dtypes: float64(1), int64(5), object(38)
memory usage: 662.5+ KB
```

Extract new data from these column into a new dataframe df, which we can continue to modify without affecting the original dataframe

```
Gender      0
Caste       0
Age_Group   0
Education   0
DistrictName 0
VillageName 0
Occupation  0
Which party has an edge in this village? 0
dtype: int64
```

we've mow cleaned up our data amd prepared for analysis. Let's take a look at sample of rows from the dataframe.

	Gender	Caste	Age_Group	Education	DistrictName	VillageName	Occupation	Which party has an edge in this village?
660	Male	Other SC	36 - 45	Below Primary	Patan	KORDA	Owner-cultivators 1-5 Acres	BJP Clear
61	Male	Other Service OBCs	26 - 35	Others	Gir Somnath	SHEDHAYA	House-wife/husband	BJP Tight
574	Male	Patel/Patidar	46 - 60	Below Primary	Bhavnagar	ANKOLALI	Other Agricultural Workers	BJP Tight
586	Blank	Prajapati/Kumbhar	46 - 60	Blank	Banaskantha	CHANGA	Blank	INC Tight
1393	Male	Rajputs	46 - 60	Below Primary	Bhavnagar	UKHRALA	Petty shopkeeper	BJP Tight
604	Female	Rajput	36 - 45	Middle Pass / Metric Fail	Anand	BANDHNI	House-wife/husband	INC Clear
1220	Male	Others	18 - 25	Middle Pass / Metric Fail	Vadodara	ATLADRA	Small businessmen	BJP Tight
1831	Male	Others	26 - 35	Middle Pass / Metric Fail	Gandhinagar	HALISA	Agricultural labourers	BJP Tight
637	Male	Other STs	46 - 60	Middle Pass / Metric Fail	Surat	VAHEVAL	Political activists	BJP Clear
1762	Male	Prajapati/Kumbhar	46 - 60	Graduate or Equivalent	Kheda	MAATAR	Small businessmen	BJP Tight

## Exploratory Analysis and Visualization

Before starting with analysis, creating demographic i.e. country,age,education,cast etc. of the respondents look like.

Let's begin with matplotlib.pyplot and seaborn

## District

Let's look for the District from which there are responses in the survey, and plot the top 10 countries with the highest number of responses.

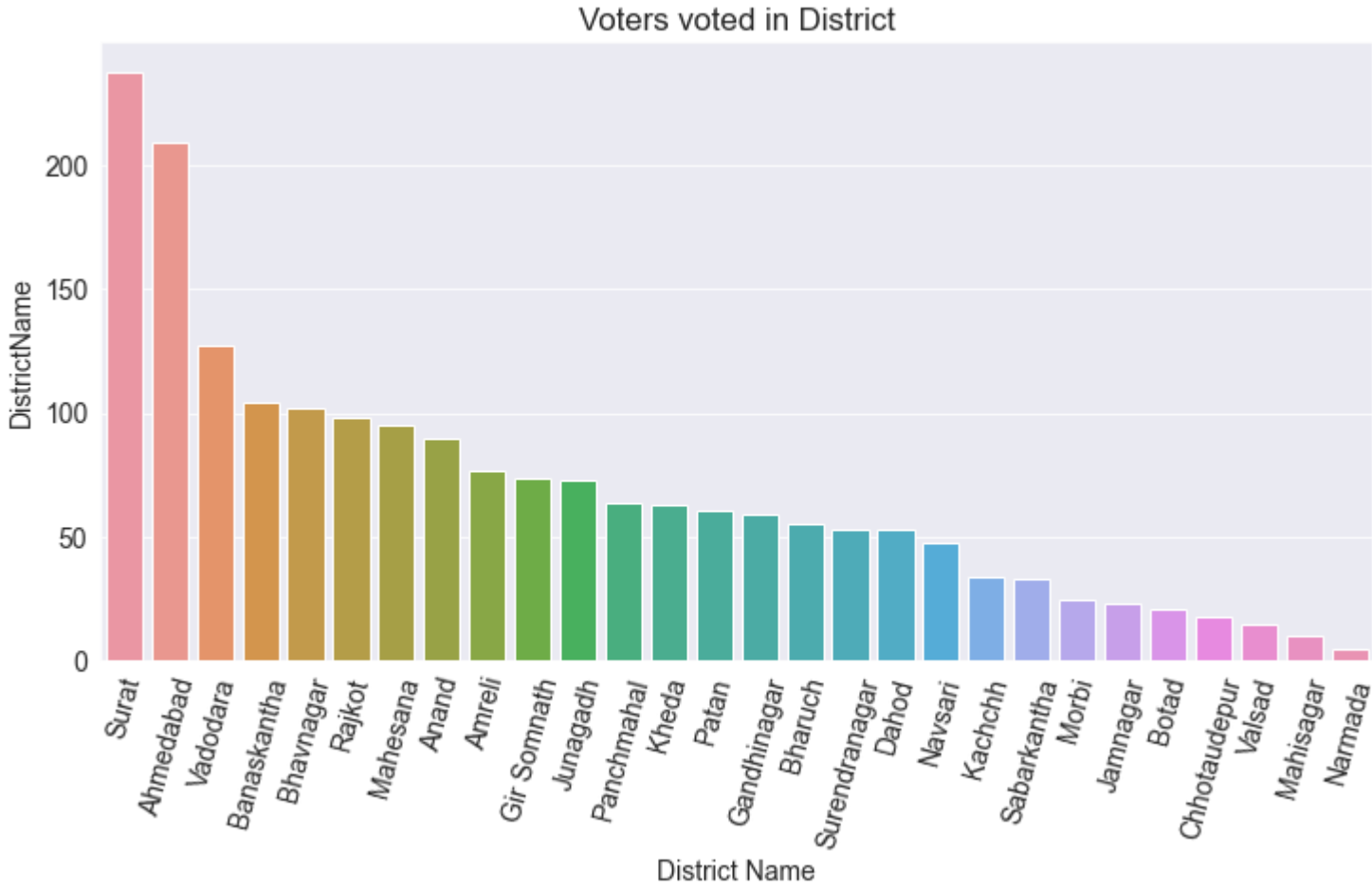
28

Let's look into District with highest number of respondents

```
Surat      238
Ahmedabad  209
Vadodara   127
Banaskantha 104
Bhavnagar  102
Rajkot     98
Mahesana   95
Anand      90
Amreli     77
Gir Somnath 74
Junagadh   73
Panchmahal 64
Kheda      63
Patan      61
Gandhinagar 59
Bharuch    55
Surendranagar 53
Dahod      53
Navsari    48
Kachchh    34
Sabarkantha 33
Morbi      25
Jamnagar   23
Botad      21
Chhotaudepur 18
Valsad     15
Mahisagar  10
```

```
Narmada      5
Name: DistrictName, dtype: int64
```

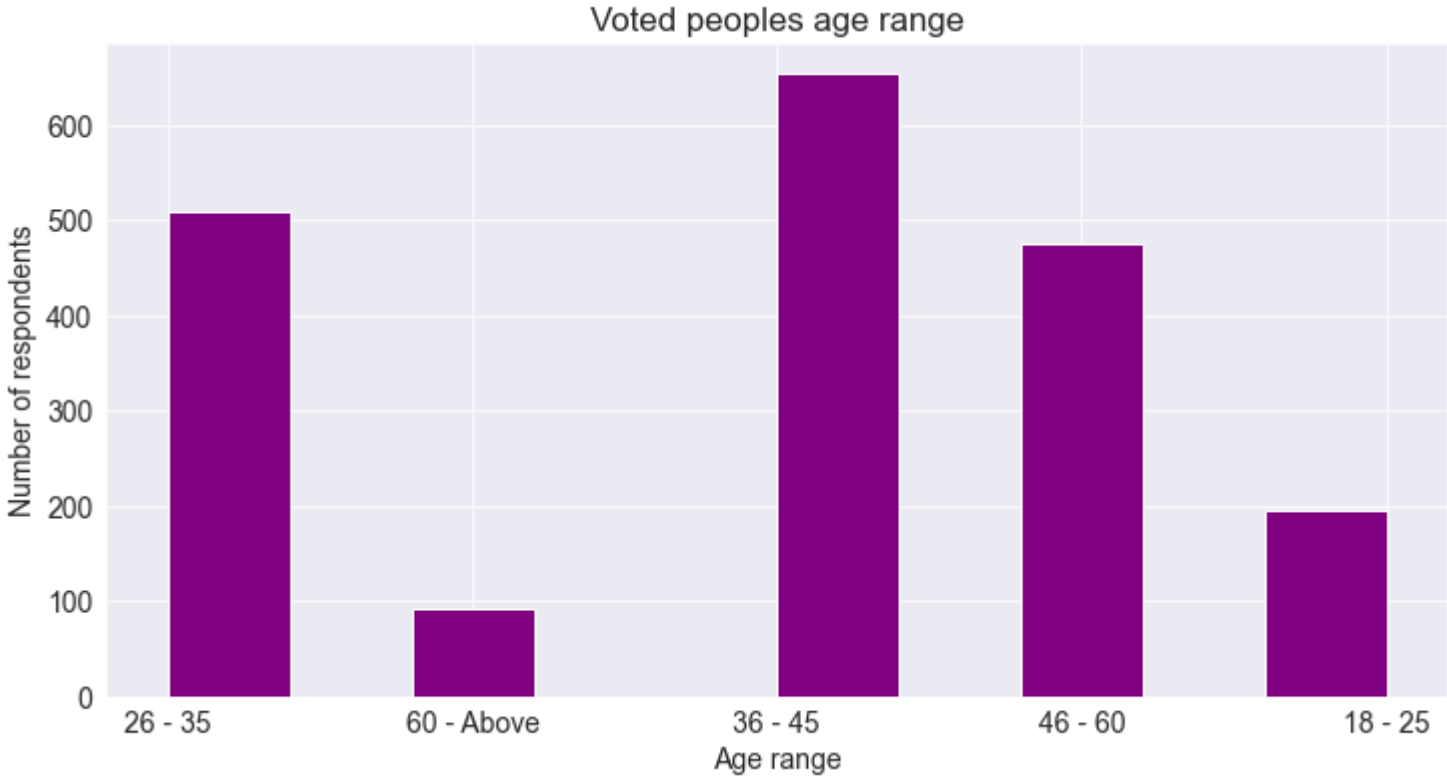
Visualizing this using barchart



It appears that a disporportionately high number of respondents are from surat and ahemdabad - which one might expect since these district have the highest populations.

## Age

The distribution of the age of respondents is also another important factor to look at, we can use a histogram to visualize it.



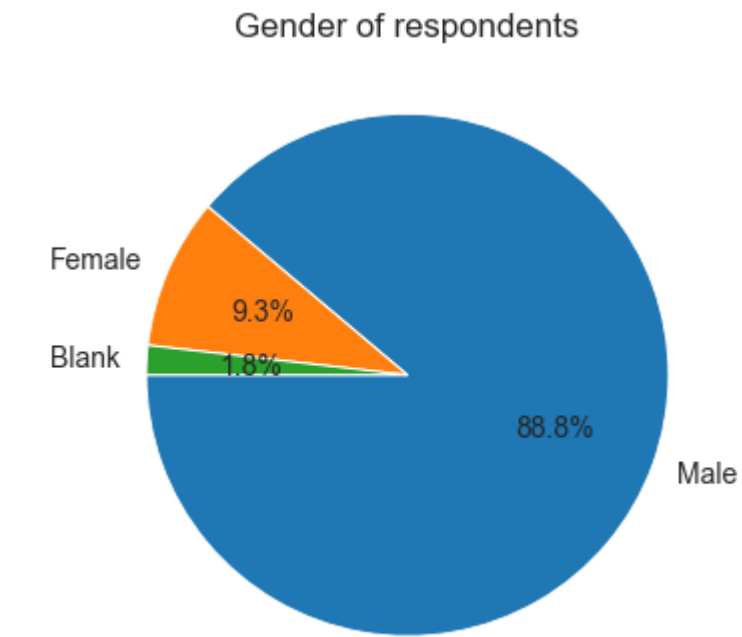
It appears that large percentage are in the age range of 36-45, 26-35 and 46-60, which is somewhat relevant as young and somewhat old peoples are interested in the political leaders while the youngest peoples are low, probably because they are busy with their life and don't much care about pliticians while 60 above peoples are not high as well, the reason might be the life span of peoples are low these times or they are too old to go anywhere and vote for someone or probably not interested.

## Gender

Let's look at the distribution of responses for the Gender.

```
Male      1712
Female    180
Blank      35
Name: Gender, dtype: int64
```

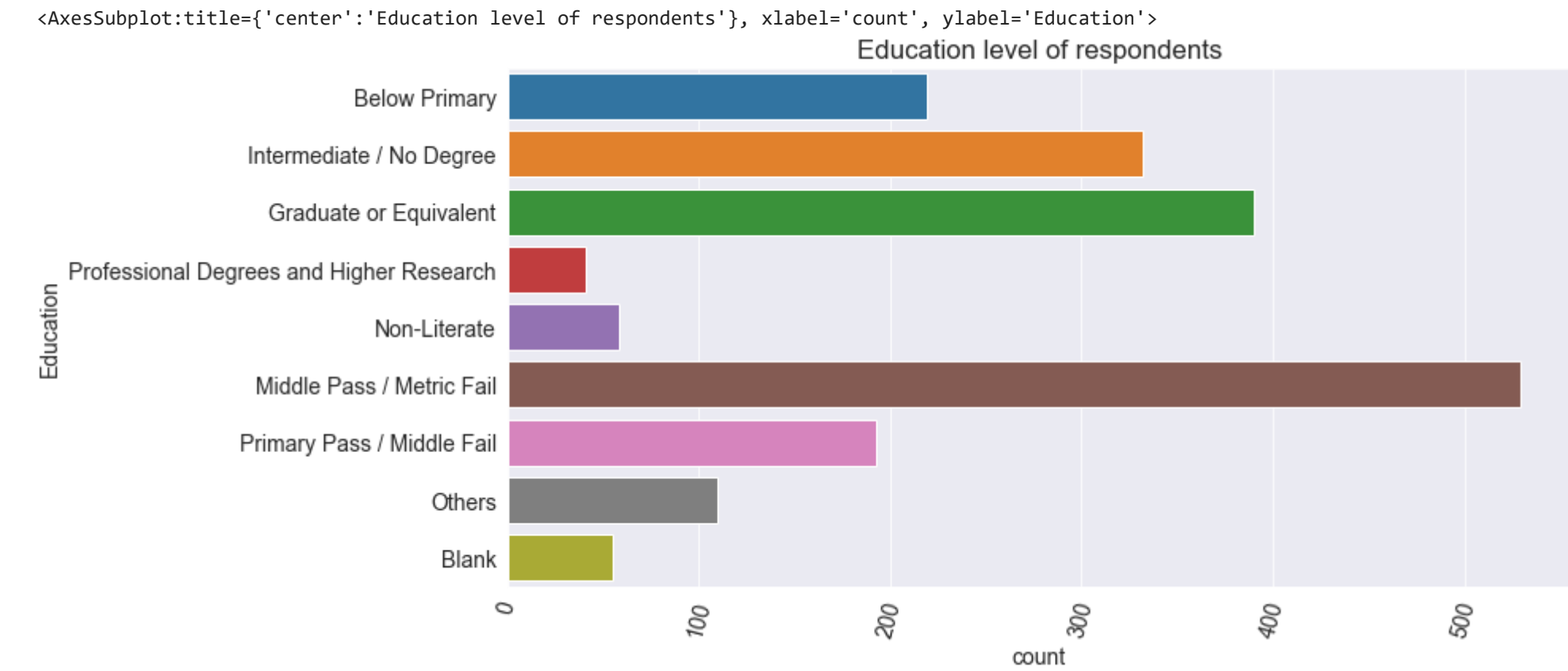
A pie chart would be a good way to visualize the distribution.



Only about 9.3% of survey respondents who have voted and the ratio of man to woman in gujrat is about 900 per 1000 males which is relatevely very low who voted. The reason might be because peoples are livint with orthodox thinking where no women are allowed to go outside and they are meant to work at home or women are not interested in politics . Literacy percentage of gujrat is around 85% including females, so the main question arises -> what is the main reason behind why females are not voting.

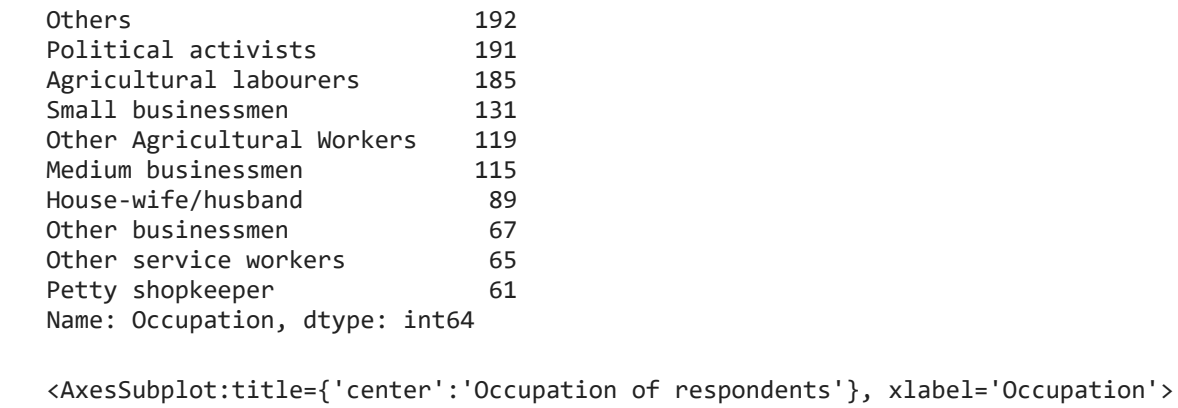
## Education Level

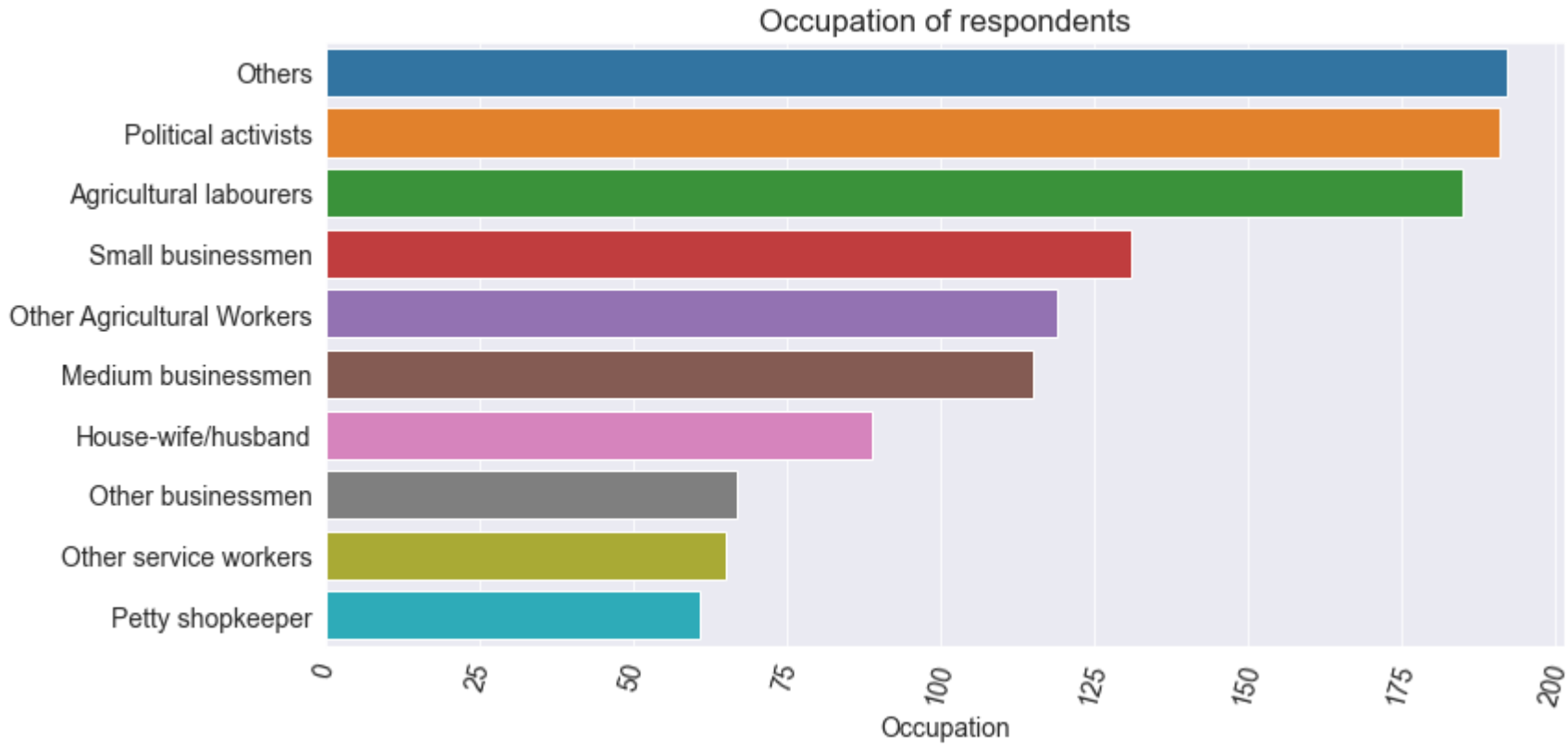
Let's compare the education levels of respondents to gain some insight into this.Let's compare the education levels of respondents to gain some insight into this.



It appears that Middle pass/Metric fail and graduate Equivalent respondents are high in number. From this we can conclude that respondents are somewhat our highly educated wile non-literate respondents are very less in number.

## Occupation





From this plot we can say that political activists and agricultural labourers are more interested in voting.

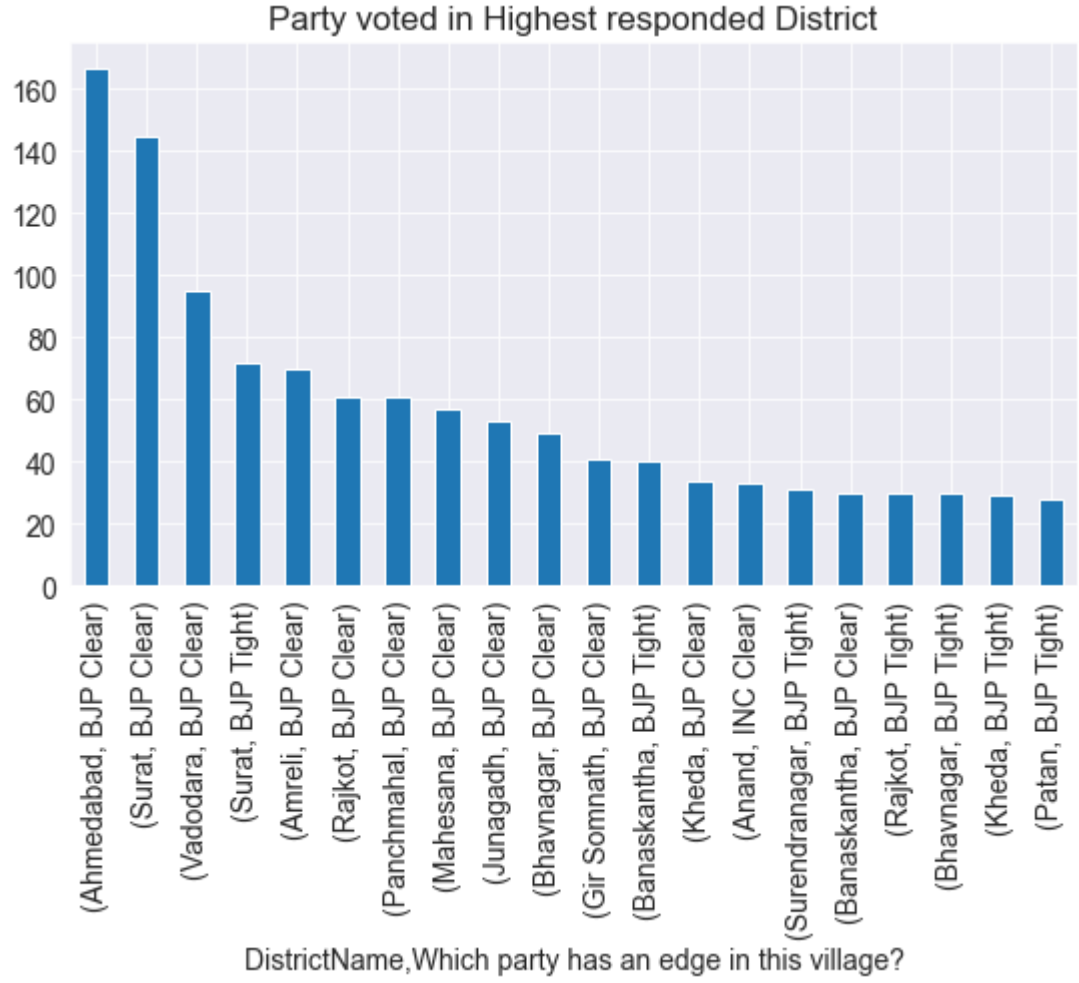
## Question and Answers

Question 1. How many respondents voted for any party from a District.

- Find the party voted in maximum responded district
- Find the party voted in minimum responded district

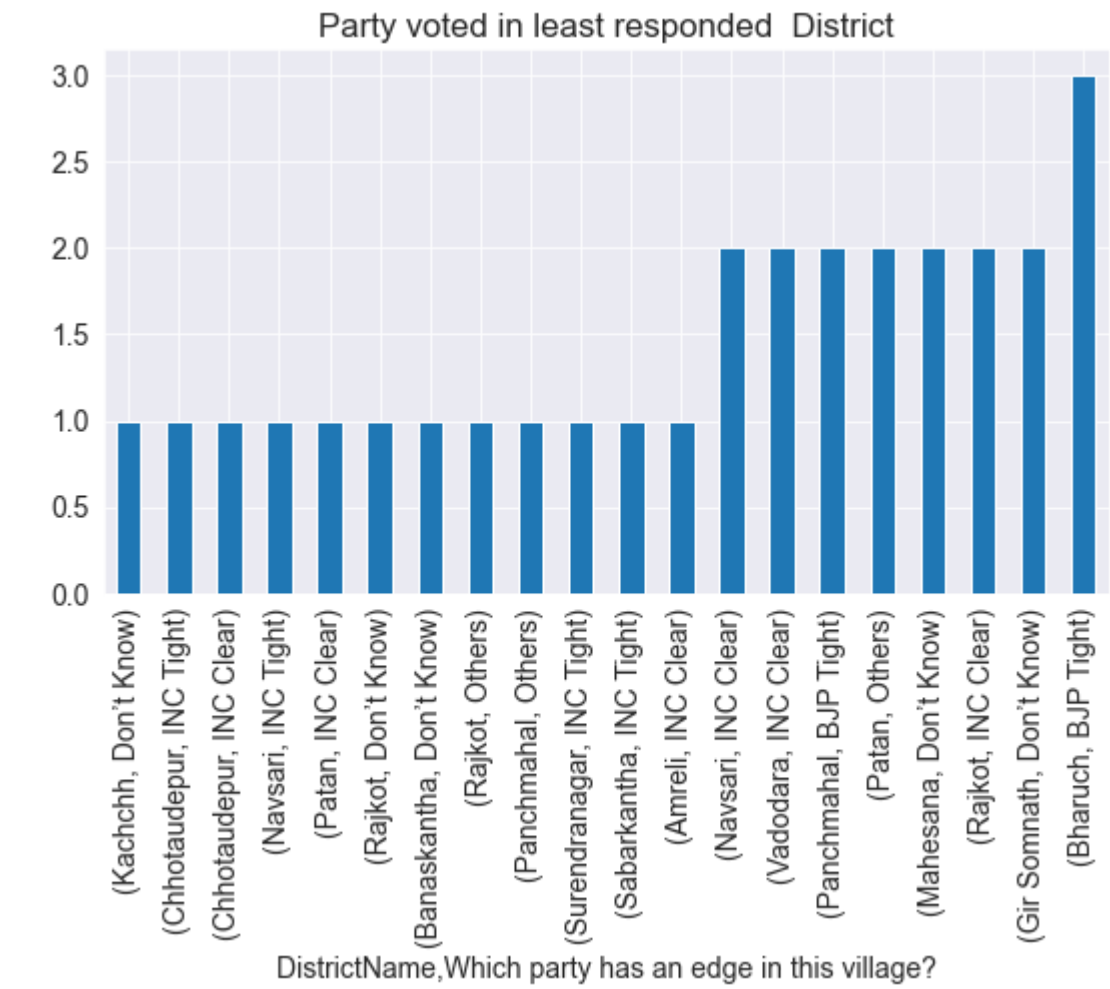
DistrictName	Which party has an edge in this village?	
Ahmedabad	BJP Clear	167
	BJP Tight	10
	INC Clear	16
	INC Tight	16
Amreli	BJP Clear	70
...		
Vadodara	INC Clear	2
	Others	3
Valsad	BJP Clear	4
	BJP Tight	3
	INC Clear	8
Length: 99, dtype: int64		

<AxesSubplot:title={'center':'Party voted in Highest responded District'}, xlabel='DistrictName,Which party has an edge in this village?'



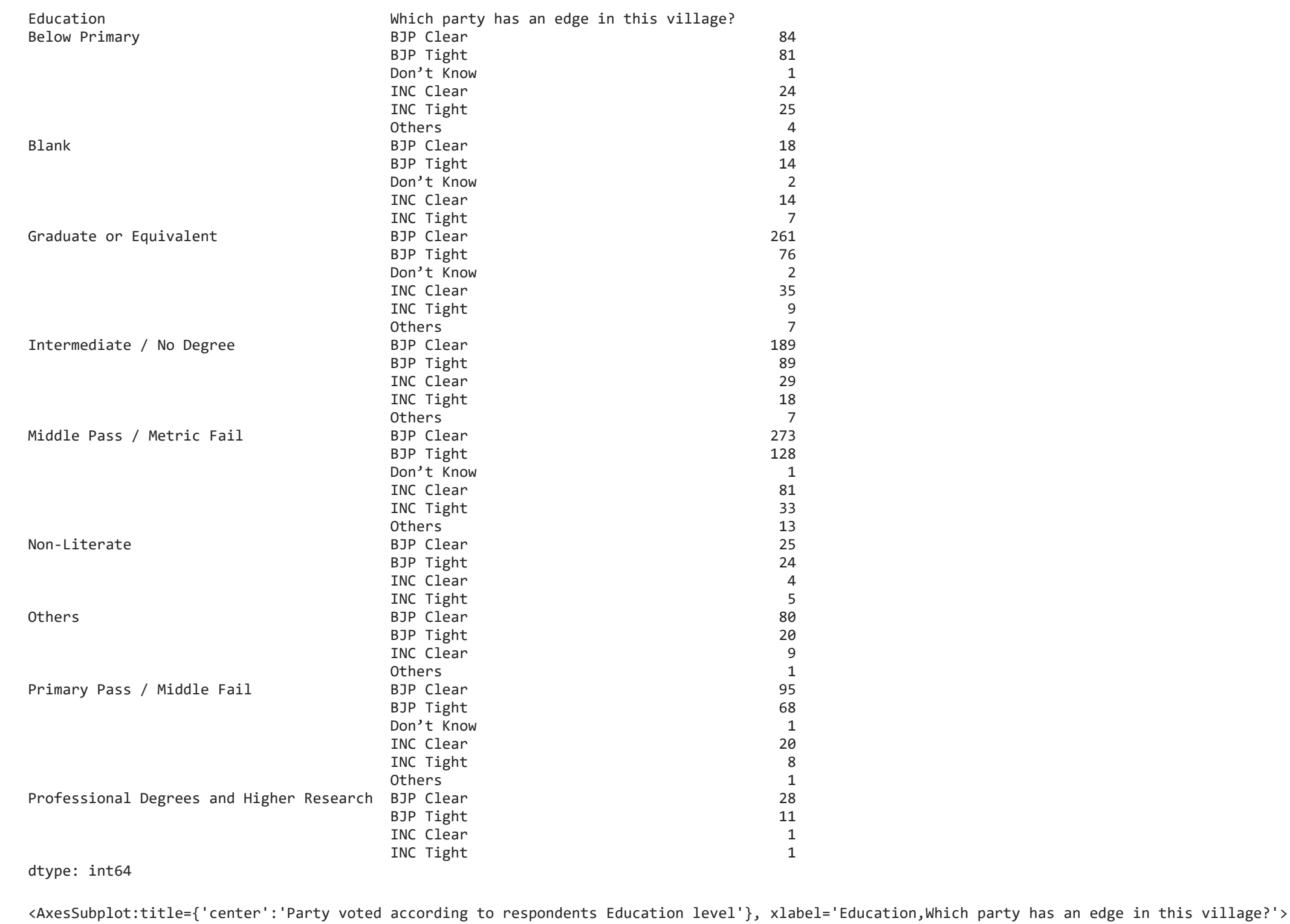
As we can see, the top populated district are being voted for BJP.

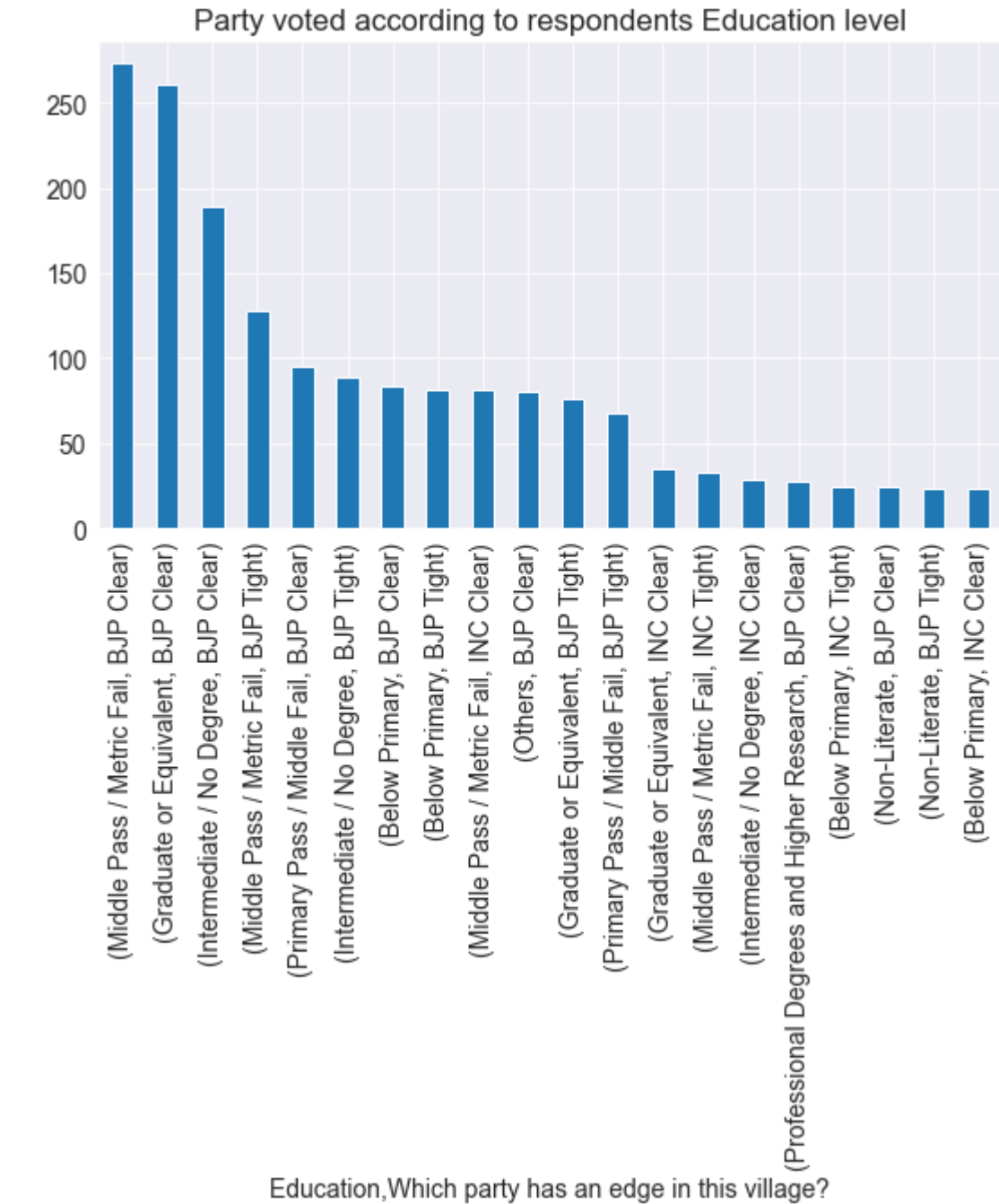
<AxesSubplot:title={'center':'Party voted in least responded District'}, xlabel='DistrictName,Which party has an edge in this village?'



As we can see, the least populated district are being voted for INC.

Question 2. Based on education level, find the maximum number of respondents who voted for any individual party.



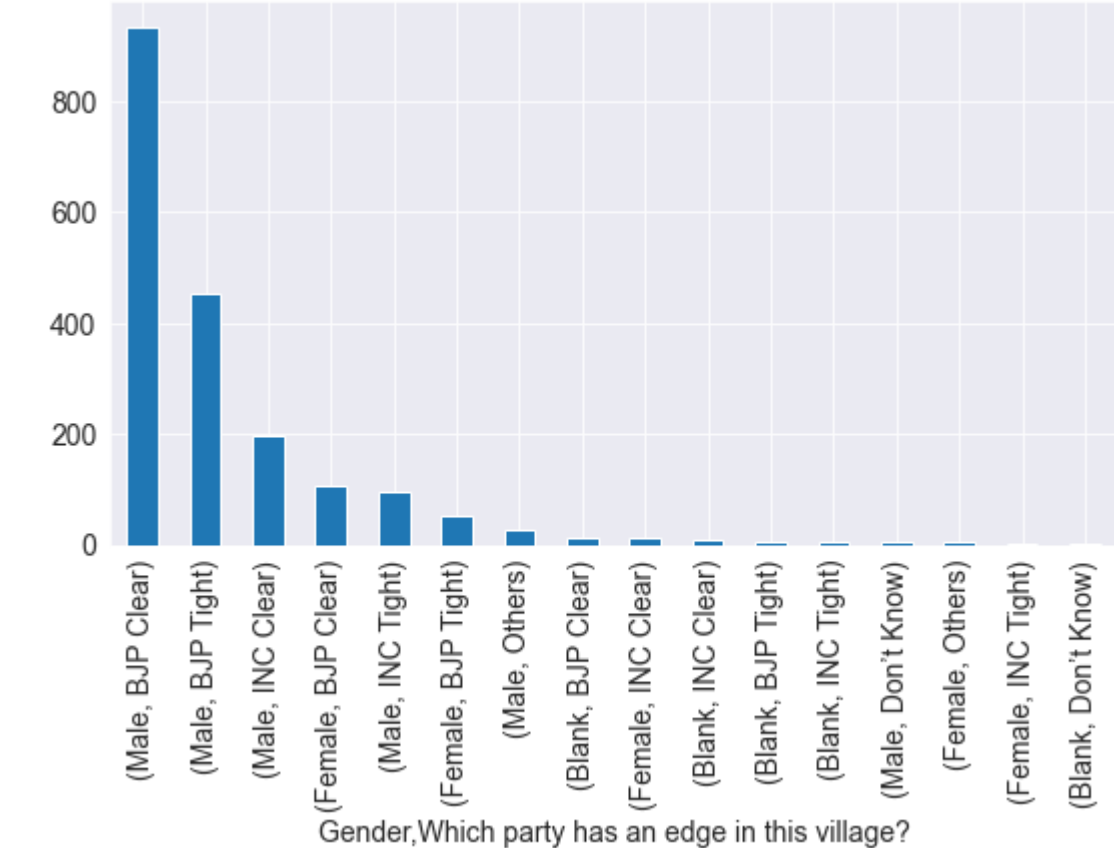


Question 3. Based on Gender, find the maximum number of respondents who voted for any individual party.

Gender	Which party has an edge in this village?	
Blank	BJP Clear	13
	BJP Tight	7
	Don't Know	1
	INC Clear	8
Female	INC Tight	6
	BJP Clear	107
	BJP Tight	53
	INC Clear	13
	INC Tight	3
	Others	4
Male	BJP Clear	933
	BJP Tight	451
	Don't Know	6
	INC Clear	196
	INC Tight	97
	Others	29

dtype: int64

<AxesSubplot:xlabel='Gender,Which party has an edge in this village?'



Maximum number of Males and females voted for BJP.

Similarly we can asnwer many more questions and visualize them