

Adithya Kiran || Chirag P Tubakad || Arpit Agarwal || Nandakrishna
PES1201700231 || PES1201700896 || PES1201701084 || PES1201701090
ASSIGNMENT 3

Problem Statement:

Perform simple Extract Transform Load (ETL) functions with Python modules on two separate datasets and merge them into one.

Dataset: 3 custom sets made from dataset of assignment 1 (uploaded to repo)

The screenshot shows a Jupyter Notebook interface with a code cell and its output. The code cell contains the following Python code:

```
input_schema={'candidate_id': 'int64', 'name': 'object', 'age': 'int64', 'constituency': 'object', 'state': 'object', 'p
print("1.csv")
display(pd.read_csv('1.csv'))
print("2.csv")
display(pd.read_csv('2.csv'))
print("3.csv")
display(pd.read_csv('3.csv'))
print("\n\n\n RESULT OF MERGING : ")
df=merge_list(['1.csv', '2.csv', '3.csv'])
display(df)
```

The output of the code is a table with 13 columns: candidate_id, pending_criminal_cases, literacy_rate, seat_share, winner, votes, voting_percentage, state, age, party, education, earnings, and constituency. The table contains 10 rows of data, with the last three rows highlighted in blue.

candidate_id	pending_criminal_cases	literacy_rate	seat_share	winner	votes	voting_percentage	state	age	party	education	earnings	constituency
4588	3	0.6650	0.031	0	314238	0.2954	TELANGANA	53	INC	12th Pass	2.119100e+07	ADILABAL
5069	52	0.6650	0.031	1	377374	0.3548	TELANGANA	52	BJP	12th Pass	2.867964e+06	ADILABAL
5519	0	0.6972	0.147	0	435329	0.3801	UTTAR PRADESH	47	BSP	Post Graduate	1.115325e+08	AGRA
5521	0	0.6972	0.147	0	45149	0.0394	UTTAR PRADESH	54	INC	Post Graduate	5.507575e+07	AGRA
5522	5	0.6972	0.147	1	646875	0.5648	UTTAR PRADESH	58	BJP	Doctorate	6.566751e+07	AGRA
5813	0	0.6650	0.031	0	4548	0.0043	TELANGANA	47	IND	Post Graduate	6.700000e+05	ADILABAL
5814	0	0.6650	0.031	0	318814	0.2997	TELANGANA	54	TRS	Post Graduate	1.847789e+07	ADILABAL
5815	0	0.6650	0.031	0	5241	0.0049	TELANGANA	27	Janasena Dethu	Post Graduate	5.000000e+04	ADILABAL

Below the table, the Jupyter Notebook shows the output of the code:

```
In [7]: all_schemas
Out[7]: {'1.csv': {'candidate_id': 'int64',
'no_pending_criminal_cases': 'int64',
'literacy_rate': 'float64',
'seat_share': 'float64',
'winner': 'int64',
'votes': 'int64',
```



