

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

```
In [2]: df=pd.read_csv("ApplicantData - ApplicantData.csv")
df
```

Out[2]:

	App_ID	Country	University	Phone_Number
0	12345	India	Illinois Institute of Technology	9823241234
1	12345	India	Illinois Institute of Technology	8805617501
2	12345	India	Illinois Institute of Technology	18019011222
3	347397	Nigeria	Illinois Institute of Technology	7738599513
4	347397	Nigeria	Illinois Institute of Technology	919182706838
...
37870	-	Bahrain	Illinois Institute of Technology	97334305522
37871	-	Bahrain	Illinois Institute of Technology	97336598776
37872	-	Bahrain	Illinois Institute of Technology	97338787272
37873	-	australia	Illinois Institute of Technology	61432347819
37874	-	Bahrain	Illinois Institute of Technology	97335911259

37875 rows × 4 columns

```
In [3]: df.shape
```

Out[3]: (37875, 4)

```
In [4]: df_encoded=df.apply(lambda col: col.astype("category")).cat.codes
```

```
In [5]: df_encoded.corr()
```

Out[5]:

	App_ID	Country	University	Phone_Number
App_ID	1.000000	-0.061496	NaN	-0.052387
Country	-0.061496	1.000000	NaN	-0.113563
University	NaN	NaN	NaN	NaN
Phone_Number	-0.052387	-0.113563	NaN	1.000000

```
In [6]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37875 entries, 0 to 37874
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   App_ID      37875 non-null   object 
 1   Country     37875 non-null   object 
 2   University   37875 non-null   object 
 3   Phone_Number 37875 non-null   object 
dtypes: object(4)
memory usage: 1.2+ MB
```

In [7]: `df.describe()`

	App_ID	Country	University	Phone_Number
count	37875	37875	37875	37875
unique	15411	856	1	18387
top		India	Illinois Institute of Technology	917506942994
freq	2676	12234	37875	47

In [8]: `df.head()`

	App_ID	Country	University	Phone_Number
0	12345	India	Illinois Institute of Technology	9823241234
1	12345	India	Illinois Institute of Technology	8805617501
2	12345	India	Illinois Institute of Technology	18019011222
3	347397	Nigeria	Illinois Institute of Technology	7738599513
4	347397	Nigeria	Illinois Institute of Technology	919182706838

In [9]: `df.tail()`

	App_ID	Country	University	Phone_Number
37870	-	Bahrain	Illinois Institute of Technology	97334305522
37871	-	Bahrain	Illinois Institute of Technology	97336598776
37872	-	Bahrain	Illinois Institute of Technology	97338787272
37873	-	australia	Illinois Institute of Technology	61432347819
37874	-	Bahrain	Illinois Institute of Technology	97335911259

In [10]: `df["App_ID"].nunique()`

Out[10]: 15411

```
In [11]: df["App_ID"].unique()
```

```
Out[11]: array(['12345', '347397', '358065', ..., '98077675', '9876789', '425358'],  
              dtype=object)
```

```
In [12]: df["App_ID"].value_counts()
```

```
Out[12]: App_ID  
`          2676  
-           558  
.....       92  
.....       88  
.....       66  
...  
345411      1  
356735      1  
465766      1  
343181      1  
425358      1  
Name: count, Length: 15411, dtype: int64
```

```
In [13]: mask = df["App_ID"].astype(str).str.isnumeric()  
print(mask.value_counts())
```

```
App_ID  
True     33119  
False    4756  
Name: count, dtype: int64
```

```
In [14]: df=df[mask].copy()  
df.reset_index(drop=True, inplace=True)
```

```
In [15]: df.shape
```

```
Out[15]: (33119, 4)
```

```
In [16]: df["App_ID"].value_counts()
```

```
Out[16]: App_ID  
418630    37  
474311    20  
451193    20  
438353    20  
478295    19  
..  
457715    1  
423011    1  
431886    1  
456664    1  
425358    1  
Name: count, Length: 15159, dtype: int64
```

```
In [17]: df["App_ID"].unique()
```

```
Out[17]: array(['12345', '347397', '358065', ..., '98077675', '9876789', '425358'],  
              dtype=object)
```

```
In [18]: df["App_ID"].value_counts()
```

```
Out[18]: App_ID  
418630    37  
474311    20  
451193    20  
438353    20  
478295    19  
      ..  
457715     1  
423011     1  
431886     1  
456664     1  
425358     1  
Name: count, Length: 15159, dtype: int64
```

```
In [19]: df["Country"].nunique()
```

```
Out[19]: 817
```

```
In [20]: df["Country"].unique()
```

```
Out[20]: array(['India', 'Nigeria', 'saarthaksingh05@gmail.com',
   'satya.sai1881@gmail.com', 'sharmaishaan16@gmail.com',
   'pillir1026@outlook.com', 'shalinidec05@gmail.com',
   'rameshpriyanka536@gmail.com', 'samjainsamj16@gmail.com',
   'ratneshry06@gmail.com', 'dhruvvekariya.vmc18@gmail.com',
   'kanishk.9871@gmail.com', 'gayathri.keenala@gmail.com',
   'lakshaymunjal17@gmail.com', 'shah.rahulsalesh@gmail.com',
   'sekhar.fall2022@gmail.com', 'glourdurani@gmail.com',
   'polishetty.jyothi08@gmail.com', 'chsrilatha.eduf23@gmail.com',
   'saimukeshjakkula2023@gmail.com', 'phanendrakatta@gmail.com',
   'garv.career@gmail.com', 'likithapalakolanu08@gmail.com',
   'saisandeep.pashem007@gmail.com', 'nehareddy6721@gmail.com',
   'polanisaibalaji@gmail.com', 'reddyabhinay037@gmail.com',
   'ganipineni.vinay999@gmail.com',
   'pasupulattepavankalyan583@gmail.com', 'ganesh.edu2023@gmail.com',
   'gauthamkanyadi@gmail.com', 'riyasubashmendon@gmail.com',
   'ravitejanuthakki123@gmail.com', 'namitanaikwadi04@gmail.com',
   'parthshroff1007@gmail.com', 'siddardh3.v@gmail.com',
   'shrushtigugale14@gmail.com', 'swapna.lbf286@gmail.com',
   'satam.arya19@gmail.com', 'Hemilacharya2023@gmail.com',
   'nsitsahilbhola@gmail.com', 'saisreekarsesham@gmail.com',
   'juturuprem2000@gmail.com', 'manobhiram.pabbathi@gmail.com', 'na',
   'naq', 'farheen8223@gmail.com', 'sriram18042002@gmail.com',
   'sonalisawant79@gmail.com', 'gowtham79789@gmail.com',
   'nrupamythali01@gmail.com', 'gpavankalyan1102@gmail.com',
   'nadellaavinash7@gmail.com', 'spkarthiek2424@gmail.com',
   'pranithavelusamysundararaj@gmail.com',
   'gopikrishna.spring2023@gmail.com', 'preethamrakshith11@gmail.com',
   'kmugdha99@gmail.com', 'samheeta.g1@gmail.com',
   'rhrithikreddy666@gmail.com', 'karthikreddychanda123@gmail.com',
   'maddy.vempaty@gmail.com', 'sasank1613@gmail.com',
   'pradeepcj98@gmail.com', 'chandana.neerubhavi@gmail.com',
   'imrannshaiks@gmail.com', 'sawantlikhitin@gmail.com',
   'kartikdhaduk98.2023@gmail.com',
   'manishankarchamidisetty@gmail.com', 'sgs24999@gmail.com',
   'desamvenkateswarareddy2762@gmail.com',
   'sakshinagapure28@gmail.com', 'harshithavenkatesh2204@gmail.com',
   'ksdchinni1784@gmail.com', 'manasgarg123@gmail.com',
   'lakshmikolasani77@gmail.com', 'bollaharika13@gmail.com',
   'sahilhambir26@gmail.com', 'sahithi.balerao29@gmail.com',
   'navanishwar123@gmail.com', 'sadwik3457@gmail.com',
   'harshavardhan564@gmail.com', 'ds4365778@gmail.com',
   'rishithareddy835@gmail.com', 'jhashasri.vjaf17@gmail.com',
   'chaithanyakumarasm19@gmail.com', 'preetham318@gmail.com',
   'prashanthreddynarahari@gmail.com',
   'shraddha.rakholiya9@gmail.com', 'kommitiakhilareddy@gmail.com',
   'kachiomkar31@gmail.com', 'nimishvaidya99@gmail.com',
   'kishoreshiva123456@gmail.com', 'harmannanot@gmail.com',
   'divya.varghese2108@gmail.com', 'mahaksharma8899@gmail.com',
   'rakshaanarayan01@gmail.com', 'mediboinamalleswari2@gmail.com',
   'jyothi290600@gmail.com', 'ndtaraka09@gmail.com',
   'ksk21082002@gmail.com', 'nakkadeepak01@gmail.com',
   'ritwikadas0707@gmail.com', 'rohit18102001@gmail.com',
   'reddyreddynayani@gmail.com',
   'durgaprasad.pasumarthi123@gmail.com',
   'kumyeshwanthreddy1999@gmail.com', 'sairamdubbaka05@gmail.com',
```

'poojithsainni2288@gmail.com', 'maniss.mishra23@gmail.com',
'harsh5432@yahoo.in', 'Nagesh.lbf475@gmail.com',
'iparkarrohit@gmail.com', 'sheetalkan1999@gmail.com',
'nidhichandarana03@gmail.com', 'mangeshkingaonkar@yahoo.com',
'rohithreddy.ywf17@gmail.com', 'pusulurivarnitha@gmail.com',
'manikanthatthapuram2000@gmail.com', 'snehshah741@gmail.com',
'sakevinay5138@gmail.com', 'hemanthk.edu23@gmail.com',
'piyushhinduja20@gmail.com', 'kadiyam jitendra@gmail.com',
'kevingodfreyverpula@gmail.com', 'nupoormathpati@gmail.com',
'dyavanavenivamshikrishna@gmail.com', 'pranavagyoti@gmail.com',
'likithalikkky.vsln@gmail.com', 'syedhashirahmed0507@gmail.com',
'rupakdgr55@gmail.com', 'jaswanth5249@gmail.com',
'rampuramakshay7@gmail.com', 'manoharb.fall2023@gmail.com',
'mailshivakumar.dsk@gmail.com', 'harshithdr10@gmail.com',
'sanjayb.af212@gmail.com', 'rishi rajani97@gmail.com',
'santhoshkn.edu2023@gmail.com', 'poojagajare30@gmail.com',
'rithvikrangaraj2@gmail.com', 'lohit.lbf620@gmail.com',
'parvathimalapati7@gmail.com', 'sundarmachani@gmail.com',
'pulimamidibhargavi1817@gmail.com', 'pradeepbandla123@gmail.com',
'mehakmathur2000@gmail.com', 'mailto.farhama@gmail.com',
'kandoiachal@gmail.com', 'sathwikthammi@gmail.com',
'ramyakosaraju0201@gmail.com', 'shreyathakur2016@gmail.com',
'haindu.15@gmail.com', 'revanthreddyk1@gmail.com',
'nikhil.yaf93@gmail.com', 'krishsakharkar03@gmail.com',
'saitejavasa2103@gmail.com', 'mosalichandhana99@gmail.com',
'gudurureddy0399@gmail.com', 'girishsaivarma@gmail.com',
'gloknadh.ao@gmail.com', 'mithali9999@gmail.com',
'Shahjaimin610@gmail.com', 'chhetamayur@gmail.com',
'shaik bajivali169@gmail.com', 'manikantadavuluri08@gmail.com',
'prathyush.lebaku7@gmail.com', 'gunsai22@gmail.com',
'dantulurimanojvarma study@gmail.com', 'mourya.eduf23@gmail.com',
'rutwik126@gmail.com', 'linson24rodrigues@gmail.com',
'prashantsalunke2001@gmail.com', 'mohdabbaskhan21@gmail.com',
'eshaan0209@gmail.com', 'krinas393@gmail.com',
'lakshmisashrika@gmail.com', 'charansaipuchapotula123@gmail.com',
'parwazekam@gmail.com', 'deepthisanikommu512@gmail.com',
'chaitanya.kirkase@gmail.com', 'mohanendraapplication@gmail.com',
'soumyaganesh298@gmail.com', 'chsktg031@gmail.com',
'shivamsatwara842@gmail.com', 'jaswanthmandava027@gmail.com',
'rohithyvcs@gmail.com', 'ravi2001raghav@gmail.com',
'parmeshwalunj@gmail.com', 'hema724821@gmail.com',
'kranthipedamaggi@gmail.com', 'kiranmayeev156@gmail.com',
'khushipatel2871@gmail.com', 'p.pranee143@gmail.com',
'rohan.ghirnikar21@gmail.co', 'neham1998@gmail.com',
'deepaktn282@gmail.com', 'sairamcharan2457@gmail.com',
'sriranjan.srikanth4@gmail.com', 'sakethreddy5880@gmail.com',
'p.varunrao123@gmail.com', 'sakhamurigeethika123@gmail.com',
'muthireddyakash@gmail.com', 'klharshitha1@gmail.com',
'contact.rahal.balaji@gmail.com', 'patelviraj1091@gmail.com',
'kenilpatelk041@gmail.com', 'rohithkandula937@gmail.com',
'sreejak.vfd271@gmail.com', 'cherukurisirishasiri@gmail.com',
'rajithamay7@gmail.com', 'nikhilsriram2001@gmail.com',
'srikup20@gmail.com', 'pavankumar98gurram@gmail.com',
'hrishi.dighe@gmail.com', 'lakshmanlucky125@gmail.com',
'manogna6864@gmail.com', 'dhanushuppala pati852@gmail.com', '- ',
'prajapatizeel240024@gmail.com', 'ronakca2023@gmail.com',

'navida08@gmail.com', 'msts.mohilk16@gmail.com',
'suthara727@gmail.com', 'prateekgupta2601@gmail.com',
'reddy.rishi2000@gmail.com', 'smitasarkar2001@gmail.com',
'kasalapoojitha.2021@gmail.com', 'miloni2023@gmail.com',
'pyash7621@gmail.com', 'hunnyjumani26@gmail.com',
'depatel.2530@gmail.com', 'harshitha.vfd295@gmail.com',
'jagadeeshutukuru@gmail.com', 'soubhagyakkenna@gmail.com',
'dharmanavasuvd@gmail.com', 'shreyauttarwar85@gmail.com',
'masumbabamohmmad123@gmail.com', 'mohammadjaveedbaig@gmail.com',
'harshitha.kcf89@gmail.com', 'shrijadhav9763@gmail.com',
'sukruthidurusoji8940@gmail.com', 'dhavalwaghare@gmail.com',
'kvinay072@gmail.com', 'maddinenilakshmiprasanna7@gmail.com',
'saiyanurag2023@gmail.com', 'shivatmika.29@gmail.com',
'kumargaurav.sinha260698@gmail.com', 'sriharshavemuri7@gmail.com',
'eerladinaveeresh@gmail.com', 'mahakpurania@gmail.com',
'manojkumarjala0098@gmail.com', 'pranjanigopani53@gmail.com',
'chaitra.gurjar@gmail.com', 'jadhabavakshay0701@gmail.com',
'sivasaiudaya@gmail.com', 'keerthi.vgf334@gmail.com',
'jvpinnamaraju@gmail.com', 'deekshith1299@gmail.com', 'Taiwan',
'Pakistan', 'ijaz.ahmed04@gmail.com', 'sarimraiz11@gmail.com',
'junaid.khan@utdallas.edu', 'javeria.ayyub@gmail.com',
'mahambasit2000@gmail.com', 'rohin.sharma.1993@gmail.com',
'jmarium20@gmail.com', 'hayatrahman731@gmail.com',
'mfuzailedu@gmail.com', 'sohaib.salahuddinakbar@gmail.com',
'omer_munam@yahoo.com', 'hasnain33920@gmail.com',
'fizanaeem0012@gmail.com', 'hiba.hasan.siddiqui@gmail.com',
'renishaacharya23@gmail.com', 'prastut79@gmail.com',
'ghimiresudeep07@gmail.com', 'pratikmb7@gmail.com',
'dinesh.745421@bkmc.tu.edu.np', 'ram.pageni@pmc.edu.np',
'sanjeevshapkota@gmail.com', 'regmi.nirmala1@gmail.com',
'ddmelamchi@gmail.com', 'samipradhan07@gmail.com',
'nshrey53@gmail.com', 'nigammalla@gmail.com', 'pasabhui@gmail.com',
'safalgautam2593@gmail.com', 'ravindrashahktm@gmail.com',
'j.tasnimkhan@outlook.com', 'sumaaairakutub@gmail.com',
'Hasan.sust.fet@gmail.com', 'nusratanjum4545@gmail.com',
'kaushikpauljoy@gmail.com', 'nazmulhaque.aiub@gmail.com',
'sharuf.sb.hossain@gmail.com', 'koushik_dey@hotmail.com',
'farzanayashmin119@gmail.com', 'ragibrownak2825@outlook.com',
'syeda.maliha.mon@gmail.com', 'sudip.kumar.sengupta@g.bracu.ac.bd',
'protiknag08@gmail.com', 'priam.me.buet.15@gmail.com',
'ipshita.ishrar@g.bracu.ac.bd', 'shahina0048@gmail.com',
'malihakabir07@gmail.com', 'safayat.b.hakim@iut-dhaka.edu',
'kawsarfakir96@gmail.com', 'Bangladesh',
'csc103.fall2010.10305038@gmail.com', 'swarajitboney@gmail.com',
'rifatrahman097@gmail.com', 'hossenms@gmail.com',
'ridon993@gmail.com', 'refatnoor.eee@gmail.com',
'ruzzaibrahim@gmail.com', 'mnamiz@gmail.com', 'mtbilto@gmail.com',
'idowuidelani@outlook.com', 'mollabakare@gmail.com',
'imamzhb157@gmail.com', 'China', 'callistusokonkwor@gmail.com',
'prakriti.pokhrel3@gmail.com', 'naseebabdulkareem@gmail.com',
'sayandeeptaity328@gmail.com', 'mpentakota01@gmail.com',
'rahul99.veeramachaneni@gmail.com', 'pranavkurundkar99@gmail.com',
'saitarun.y.vjaf84@gmail.com', 'rithwikreddye@gmail.com',
'navyasribommakanti8@gmail.com', 'meghana.viha@gmail.com',
'krithikareddy382@gmail.com', 'pavan.adusumalli456@gmail.com',
'pavanmondeddu@gmail.com', 'srk.nlpt@gmail.com',

'hanimi.uf76@gmail.com', 'sanskriteerajpal@gmail.com',
'suryatej.duppanapudi@gmail.com', 'lahari.lbf307@gmail.com',
'sudduv232@gmail.com', 'kanand00@hotmail.com',
'indrajeetkumar061096@gmail.com', 'parthnshinde101@gmail.com',
'harshaljorwekar678@gmail.com', 'syedarfath2505@gmail.com',
'dangichure17@gmail.com', 'muzaich78@gmail.com',
'koechabby2021@gmail.com', 'kimathi.antonio@gmail.com',
'mzileni55@gmail.com', 'mobinanaseri9977@gmail.com',
's.saedpana@gmail.com', 'nezhlamadabadi@gmail.com',
'ehsan.nicknejad@gmail.com', 'rahim.afshaar@gmail.com',
'hosseinzamani1995@gmail.com', 'mnikoop@g.clemson.edu',
'sepideh.marjae@gmail.com', 'meysamsahebikhirkhani@gmail.com',
'erfanmirhaji@gmail.com', 'nima.sakhaee@gmail.com',
'jafaripourparvaneh@gmail.com', 'dorsar33@yahoo.com',
'kamand.taherkhani@gmail.com', 'mohammadtaghirayati@gmail.com',
'cfd.dkk@gmail.com', 'emadthegreat1@gmail.com',
'mr.pakdel.m@gmail.com', 'masi.dashtban@gmail.com',
'mahya.asadalizadeh77@gmail.com', 'nargesshaker229@gmail.com',
'marziejanbazi1999@gmail.com', 'kianakazari80@gmail.com',
'kosaraminj@gmail.com', 'farbod.naderi.nd@gmail.com',
'shahidi.negi@gmail.com', 'solbizahra94@gmail.com',
'em.mojav@gmail.com', 'jarrahiarmin@yahoo.com',
'mj.ahmadimalayeri@gmail.com', 'pedrama8825@gmail.com',
'leila.rostamian46@gmail.com', 'n.yousefi.777@gmail.com',
'panteasaber191@gmail.com', 'reza.zakerian@slu.edu',
'saeedi.sanam@gmail.com', 'sjmona99@gmail.com',
'sahab.taali@ymail.com', 'omidjaliliandb@gmail.com',
'mehrdad.tamiji@gmail.com', 'mosayebpooladi.og@gmail.com',
'ghanaatian.amirhossein7@gmail.com', 'f.darehgazani@gmail.com',
'servat.mahnaz@gmail.com', 'rozaghaedi90@gmail.com',
'neda.saffari7115@gmail.com', 'hw.tab.au@gmail.com',
'pkisara28@gmail.com', 'rshahbazimehrizi@hawk.iit.edu',
'farnazrahmi@gmail.com', 'sakshikannav@gmail.com',
'bommena.haswika@gmail.com', 'gnishanth28600@gmail.com',
'ravi.lns400@gmail.com', 'sahithivangala23@gmail.com',
'sridev678@gmail.com', 'nitya2411@gmail.com',
'sriramchowdary936@gmail.com', 'jayaram.ywf60@gmail.com',
'spandana.y.vjaf32@gmail.com',
'srinagavenkatavijaykumarp@gmail.com', 'sushantbodke77@gmail.com',
'rishihasharma1@gmail.com', 'noyaljosephbinu98@gmail.com',
'chougule.omkar10@gmail.com', 'mohdhasanf06@gmail.com',
'rl558609@gmail.com', 'hrithik.amf308@gmail.com',
'pratik.kht2001@gmail.com', 'naveen12051999@gmail.com',
'Navyachenna54@gmail.com', 'srikanths.yad1@gmail.com',
'pampadevi89@gmail.com', 'souvikkundu009@gmail.com',
'ramyagundapuneni.lns203@gmail.com',
'keesariupender2000@gmail.com', 'mistry.khushank@gmail.com',
'deekshikolla4@gmail.com', 'madhuri.yf141@gmail.com',
'suhass7187@gmail.com', 'gonsalvesmikhail@gmail.com',
'palaksharaskapoortwala@gmail.com', 'ganashubs98@gmail.com',
'darshanshetty183@gmail.com', 'ninadmalis3@gmail.com',
'sdharmil67@gmail.com', 'fatemehhaji1998@gmail.com',
'leila.mousazadeh.biotechnology@gmail.com',
'sadaf.bahramirad@yahoo.com', 'Mobina.kashaniyan@yahoo.com',
'neda.momeni.art@gmail.com', 'marzieh.s.forootan@gmail.com',
'kimya.yadollahpour@gmail.com', 'shivanifarahnaz@gmail.com',

'shubhamkulhariya@gmail.com', 'savanam.reddy10@gmail.com',
'rushimaadhav1999@gmail.com', 'k.rahmaditia@gmail.com',
'suldywatira@gmail.com', 'mhmdsalameen@gmail.com',
'bsjadallah@gmail.com', 'mathiasyeboah18@gmail.com',
'eoeffah001@outlook.com', 'mawufemorjnr@gmail.com',
'giacomo.terragni@gmail.com', 'lreklaoui@gmail.com',
'ke_shav@hotmail.co.uk', 'mimintambo@gmail.com',
'jnurmetova92@gmail.com', 'maureenmills77@yahoo.co.uk',
'dewantanvir@iut-dhaka.edu', 'isamuradly@gmail.com',
'South Africa', 'pavani.kcf812@gmail.com',
'dwarakeshreddy5@gmail.com', 'sahilbhaware01@gmail.com',
'jainrashmi030@gmail.com', 'duddukurulokesh4118@gmail.com',
'shivanandreddygundu@gmail.com', 'gurunath.reddy0381@gmail.com',
'garimasharmaa23@gmail.com', 'kudayakrishna10@gmail.com',
'shettyshreya.sj@gmail.com', 'himavardhan292@gmail.com',
'prajodhpragaths@gmail.com', 'suryavegesna27@gmail.com',
'shelkevaishnavi.1999@gmail.com', 'krishna1899prasad@gmail.com',
'krishna.dasyam06@gmail.com', 'sakshimpandit@gmail.com',
'kancharlasaisivani@gmail.com', 'muzaffarali.wrf175@gmail.com',
'laharibasavaraju01@gmail.com', 'rideepmoran@gmail.com',
'jahnvidesai312001@gmail.com', 'sonureddy15316@gmail.com',
'hardikkikavat01@gmail.com', 'mohammadtameemattar@gmail.com',
'shreyagupta3004@gmail.com', 'priyadarshanbeepangil@gmail.com',
'bsr1212000@gmail.com', 'neha.lbf367@gmail.com',
'pradnyarghadge@gmail.com', 'krutraiyani2001@gmail.com',
'harshareddyhvr630@gmail.com', 'manukondasagarbabu@gmail.com',
'nellurinagasaikrishna@gmail.com', 'hirtikamirghani30@gmail.com',
'jyoshnakoppala55@gmail.com', 'ssneha.bsmanian@gmail.com',
'swapnils0098@gmail.com', 'kotlaashishreddy18@gmail.com',
'saibharath.lattupalli@gmail.com', 'sangati.shah@gmail.com',
'navuluriyaswanthgufus2023@gmail.com', 'sainikhil5959@gmail.com',
'kshiranazanwar@gmail.com', 'kirthibisoi.1210@gmail.com',
'saisandhiptha.ywf58@gmail.com', 'shubhamsamanta1@gmail.com',
'dkiran100@gmail.com', 'shwettashshinde@gmail.com',
'shahjuhi481@gmail.com', 'laxminarsimha.yf50@gmail.com',
'shethiameet1@gmail.com', 'dharmikainturu@gmail.com',
'shanmukha.fall2023@gmail.com', 'srivatsan.ranganath@gmail.com',
'shyam.fall2023@gmail.com', 'poornachand778@gmail.com',
'shreegowri1234@gmail.com', 'sujan0263@gmail.com',
'galipallykushal@gmail.com', 'navanethmaadhav@gmail.com',
'charishmanadavala@gmail.com', 'shivanireddy1951@gmail.com',
'shreekms23@gmail.com', 'chintalurianirudh@gmail.com',
'mrunalipuralkar31@gmail.com', 'sphoorthysrirama28@gmail.com',
'reddyjanu11@gmail.com', 'rangisettisaisaran@gmail.com',
'harshita.af1@gmail.com', 'phanendharreddyk@gmail.com',
'koppinenitejanaidu@gmail.com',
'he got admitted into another university', 'surlekarh@gmail.com',
'shirisha2398@gmail.com', 'kandregulapraneet18@gmail.com',
'chandrashekhar199825@gmail.com', 'syam.edu2023@gmail.com',
'dnyanadajdeo@gmail.com', 'kcsanjana@gmail.com',
'sumanthpulikonda007@gmail.com', 'jahnavi.fall2023@gmail.com',
'manishasure18@gmail.com', 'maithililingala.info@gmail.com',
'caojinlucky@gmail.com', 'sismonica@naver.com',
'fangyuting1110@outlook.com', 'genie_cheng@outlook.com',
'harawoo@gmail.com', 'sleyang@unc.edu',
'freeman123459@outlook.com', 'shiqiuchen22@gmail.com',

'rooooooy_10969@163.com', 'chenyujian2023@outlook.com',
'chengkunlyu8866@gmail.com', 'jiaxinmu0@gmail.com',
'qibuswork@gmail.com', 'caoshijia0702@qq.com',
'congguocn@outlook.com', 'cirillechen002@gmail.com',
'cclu1031@yahoo.com.tw', 'dsgmsr@163.com', 'f1730116682@gmail.com',
'okohf304@gmail.com', 'linwang8989@gmail.com', 'huty.nj@gmail.com',
'serinaqin2023fall@gmail.com', 'shanwatche@gmail.com',
'chicharlieminguk@gmail.com', 'giaquang.education@gmail.com',
'poomchantara@gmail.com', 'buddhikaalahakoon94@gmail.com',
'mshammari30@gmail.com', 'kassmtakayo@yahoo.com',
'izava_a92@yahoo.com', 'jonperezbozal@gmail.com',
'miguelgranderaez@gmail.com', 'fabianamikhail@gmail.com',
'mbowaanna@outlook.com', 'smaripala@hawk.iit.edu',
'rbagul@hawk.iit.edu', 'rkessireddy@hawk.iit.edu',
'ksomisetti@hawk.iit.edu', 'fkhatri@hawk.iit.edu',
'nahuja4@hawk.iit.edu', 'rakula5@hawk.iit.edu',
'hpuppala@hawk.iit.edu', 'sgadiraju@hawk.iit.edu',
'jpuystake@hawk.iit.edu', 'nganga@hawk.iit.edu',
'sgolla1@hawk.iit.edu', 'rpayeli@hawk.iit.edu',
'sgupta97@hawk.iit.edu', 'sthummalapally@hawk.iit.edu',
'dgangappa@hawk.iit.edu', 'dchittariamaravathi@hawk.iit.edu',
'rbolla@hawk.iit.edu', 'schanamolu2@hawk.iit.edu',
'rdaggula@hawk.iit.edu', 'kkapoor5@hawk.iit.edu',
'hkhilawala@hawk.iit.edu', 'dpadte@hawk.iit.edu',
'drachakonda@hawk.iit.edu', 'mbhagalakshmi@hawk.iit.edu',
'mhonnenahallilakshmi@hawk.iit.edu', 'htejavath@hawk.iit.edu',
'qabbas@hawk.iit.edu', 'fnaeem1@hawk.iit.edu',
'masif3@hawk.iit.edu', 'mashraf3@hawk.iit.edu',
'eemmanuel@hawk.iit.edu', 'mabbasi5@hawk.iit.edu',
'hsaeed3@hawk.iit.edu', 'ftanveer@hawk.iit.edu',
'rnethi@hawk.iit.edu', 'mchaganti@hawk.iit.edu',
'npattem@hawk.iit.edu', 'rgajavalli@hawk.iit.edu',
'hningappakondakundi@hawk.iit.edu', 'ssundaram@hawk.iit.edu',
'syerramsetti@hawk.iit.edu',
'Not attending Illinois Tech, going to a higher ranked university',
'Afghanistan', 'Azerbaijan', 'Algeria', 'Armenia', 'Bahrain',
'Bolivia', 'Botswana', 'Brazil', 'Cameroon', 'Burundi',
'Congo, the Democratic Republic of the', 'Djibouti', 'Egypt',
'Eritrea', 'Estonia', 'Ethiopia', 'Gambia the', 'Ghana',
'Argentina', 'Belarus', 'Finland', 'Germany', 'Canada', 'Iran',
'Nepal', 'Myanmar', 'Kenya', 'South Korea', 'Ecuador', 'France',
'Yemen', 'Zambia', 'Zimbabwe', 'Sri Lanka', 'Rwanda',
'Saudi Arabia',
'student wants to change the degree from Doctor of Philosophy to Masters',
'not able to provide official transcript', 'Indonesia', 'Turkey',
'Australia', 'Lebanon', 'Kuwait', 'Malawi', 'Brasil', 'Colombia',
'Ethiopia, Italy', 'Ethiopia , Italy', "Cote d'Ivoire", 'Angola',
'Côte d'Ivoire', 'Cyprus', 'Dominica', 'Czech Republic',
'United Arab Emirates', 'Uganda', 'Sri lanka', 'Spain',
'Sierra Leone', 'Somalia', 'Tanzania', 'Finance', 'Gambia',
'Indonesia', 'Italy', 'Israel', 'Jordan', 'Kazakhstan', 'Mongolia',
'Panama', 'Uzbekistan', 'Tajikistan', 'Syria', 'Sweden', 'Norway',
'Swaziland', 'Russian Federation', 'Sudan', 'Latvia', 'Morocco',
'Georgia', 'Nigeria', 'Mozambique', 'Liberia', 'Malaysia',
'Hong Kong', 'Vietnam', 'Honduras', 'Iceland', 'Guyana', 'Kenya',
'Japan', 'Nepal', 'South Sudan', 'Turkmenistan', 'United Kingdom',

```
'United States of America', 'Poland', 'Switzerland', 'Singapore',
'Qatar', 'RUssia', 'Russia', 'Philippines', 'Portugal',
'Palestine', 'Netherlands', 'Netherland', 'Myanmmar', 'Thailand',
'Libya', 'Iraq', 'Senegal', 'Paraguay', 'Mexico', 'Namibia',
'Ireland', 'Lithuania', 'Seychelles', 'Republic of Korea',
'Nigeira', 'Niue', 'personal reason', 'Mauritius', 'Mororcco',
'Chile', 'nigeria', 'Lesotho', 'Oman', 'nepal', 'ethiopia', 'hk',
'HK', 'ghana', 'Ukraine', 'gambia', 'kenya', 'saudi arabia',
'nigera', 'malawi', 'albania', 'jordan', 'grenada', 'canada',
'brazil', 'china', 'chile', 'spain', 'BULGARIA', 'gabon',
'sri lanka', 'liberia', 'singapore', 'ukraine', 'columbia',
'mexico', 'Tunisia', 'kazakhstan', 'goban', 'Jordon', 'Jamaica',
'jamica', 'Peru', 'rwanda', 'ubekistan', 'zambia', 'somalia',
'south sudan', 'libya', 'malaysia', 'france', 'cameroon', 'Gabon',
'Albania', 'sudan', 'myanmar', 'cote d ivoire', 'lebanon',
'uzbekistan', 'sierra leone', 'egypt', 'congo republic', 'Congo',
'Togo', 'peru', 'benin', 'georgia', 'armenia', 'Saudi arabia'],
dtype=object)
```

In [21]: `df["Country"].value_counts()`

Out[21]: Country

Country	count
India	11366
-	5623
Ghana	3689
Nigeria	2982
Pakistan	1377
...	
garimasharmaa23@gmail.com	1
gurunath.reddy0381@gmail.com	1
shivanandreddygundu@gmail.com	1
dudukurulokesh4118@gmail.com	1
Saudi arabia	1

Name: count, Length: 817, dtype: int64

In [22]: `df = df[~df["Country"].str.contains("@", na=True)].reset_index(drop=True)`

In [23]: `df = df[df["Country"] != "-"].reset_index(drop=True)`

In [24]: `df["Country"].value_counts()`

Out[24]: Country

Country	count
India	11366
Ghana	3689
Nigeria	2982
Pakistan	1377
Iran	1275
...	
Myanmmar	1
RUssia	1
Switzerland	1
Nepaal	1
Saudi arabia	1

Name: count, Length: 200, dtype: int64

```
In [25]: df["Country"].unique()
```

```
Out[25]: array(['India', 'Nigeria', 'na', 'naq', 'Taiwan', 'Pakistan',
   'Bangladesh', 'China', 'South Africa',
   'he got admitted into another university',
   'Not attending Illinois Tech, going to a higher ranked university',
   'Afghanistan', 'Azerbaijan', 'Algeria', 'Armenia', 'Bahrain',
   'Bolivia', 'Botswana', 'Brazil', 'Cameroon', 'Burundi',
   'Congo, the Democratic Republic of the', 'Djibouti', 'Egypt',
   'Eritrea', 'Estonia', 'Ethiopia', 'Gambia the', 'Ghana',
   'Argentina', 'Belarus', 'Finland', 'Germany', 'Canada', 'Iran',
   'Nepal', 'Myanmar', 'Kenya', 'South Korea', 'Ecuador', 'France',
   'Yemen', 'Zambia', 'Zimbabwe', 'Sri Lanka', 'Rwanda',
   'Saudi Arabia',
   'student wants to change the degree from Doctor of Philosophy to Masters',
   'not able to provide official transcript', 'Indonesia', 'Turkey',
   'Australia', 'Lebanon', 'Kuwait', 'Malawi', 'Brasil', 'Colombia',
   'Ethiopia, Italy', 'Ethiopia , Italy', "Cote d'Ivoire", 'Angola',
   'Côte d'Ivoire', 'Cyprus', 'Dominica', 'Czech Republic',
   'United Arab Emirates', 'Uganda', 'Sri lanka', 'Spain',
   'Sierra Leone', 'Somalia', 'Tanzania', 'Finance', 'Gambia',
   'Indonesia', 'Italy', 'Israel', 'Jordan', 'Kazakhstan', 'Mongolia',
   'Panama', 'Uzbekistan', 'Tajikistan', 'Syria', 'Sweden', 'Norway',
   'Swaziland', 'Russian Federation', 'Sudan', 'Latvia', 'Morocco',
   'Georgia', 'NIgeria', 'Mozambique', 'Liberia', 'Malaysia',
   'Hong Kong', 'Vietnam', 'Honduras', 'Iceland', 'Guyana', 'KENya',
   'Japan', 'Nepaal', 'South Sudan', 'Turkmenistan', 'United Kingdom',
   'United States of America', 'Poland', 'Switzerland', 'Singapore',
   'Qatar', 'RUssia', 'Russia', 'Philippines', 'Portugal',
   'Palestine', 'Netherlands', 'Netherland', 'Myanmmar', 'Thailand',
   'Libya', 'Iraq', 'Senegal', 'Paraguay', 'Mexico', 'Namibia',
   'Ireland', 'Lithuania', 'Seychelles', 'Republic of Korea',
   'Nigeira', 'Niue', 'personal reason', 'Mauritius', 'Mororcco',
   'Chile', 'nigeria', 'Lesotho', 'Oman', 'nepal', 'ethiopia', 'hk',
   'HK', 'ghana', 'Ukraine', 'gambia', 'kenya', 'saudi arabia',
   'nigera', 'malawi', 'albania', 'jordan', 'grenada', 'canada',
   'brazil', 'china', 'chile', 'spain', 'BULGARIA', 'gabon',
   'sri lanka', 'liberia', 'singapore', 'ukraine', 'columbia',
   'mexico', 'Tunisia', 'kazakhstan', 'goban', 'Jordon', 'Jamaica',
   'jamica', 'Peru', 'rwanda', 'ubekistan', 'zambia', 'somalia',
   'south sudan', 'libya', 'malaysia', 'france', 'cameroon', 'Gabon',
   'Albania', 'sudan', 'myanmar', 'cote d ivoire', 'lebanon',
   'uzbekistan', 'sierra leone', 'egypt', 'congo republic', 'Congo',
   'Togo', 'peru', 'benin', 'georgia', 'armenia', 'Saudi arabia'],
  dtype=object)
```

```
In [26]: df
```

Out[26]:

	App_ID	Country	University	Phone_Number
0	12345	India	Illinois Institute of Technology	9823241234
1	12345	India	Illinois Institute of Technology	8805617501
2	12345	India	Illinois Institute of Technology	18019011222
3	347397	Nigeria	Illinois Institute of Technology	7738599513
4	347397	Nigeria	Illinois Institute of Technology	919182706838
...
26674	473550	ghana	Illinois Institute of Technology	233249887064
26675	474311	India	Illinois Institute of Technology	919440241515
26676	476192	India	Illinois Institute of Technology	918500100954
26677	477028	nigeria	Illinois Institute of Technology	2348160608306
26678	476184	nigeria	Illinois Institute of Technology	2347064846001

26679 rows × 4 columns

In [27]: df["University"].nunique()

Out[27]: 1

In [28]: df["Phone_Number"].value_counts()

Out[28]:

Phone_Number	count
233540454532	23
919346479723	22
919515029355	21
919440241515	21
2348160608306	21
..	
989123578462	1
919877692609	1
918270650006	1
918647355598	1
+234 5467121496	1

Name: count, Length: 12466, dtype: int64

In [29]: df.isnull().any()

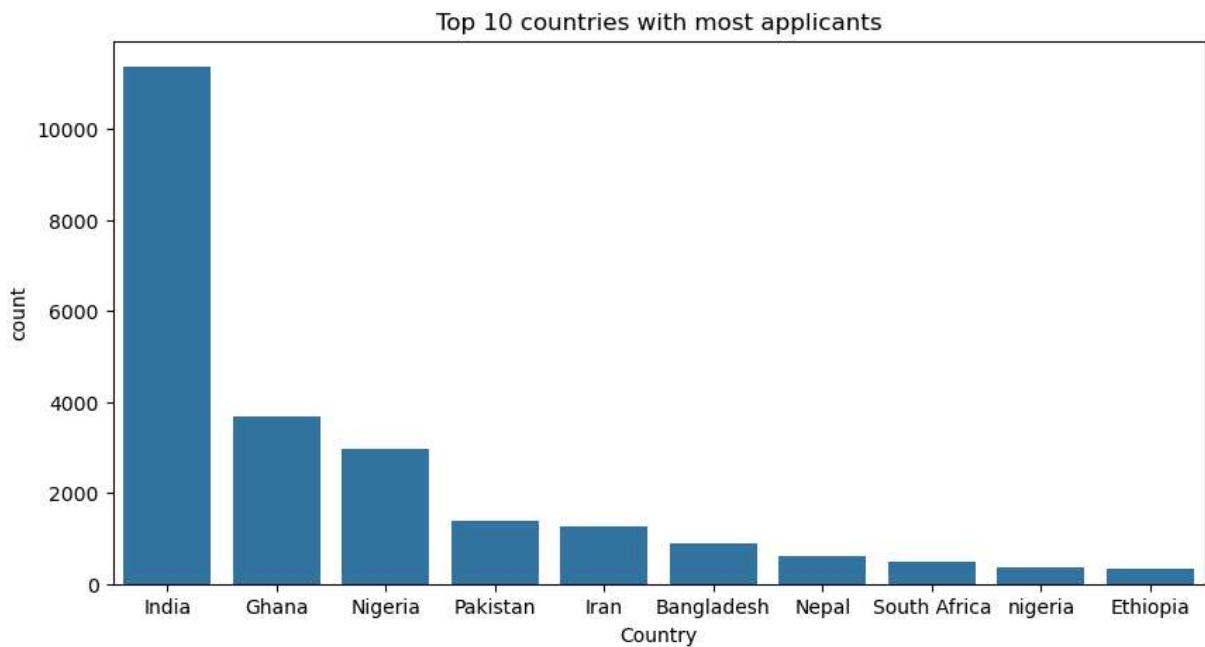
Out[29]:

App_ID	False
Country	False
University	False
Phone_Number	False

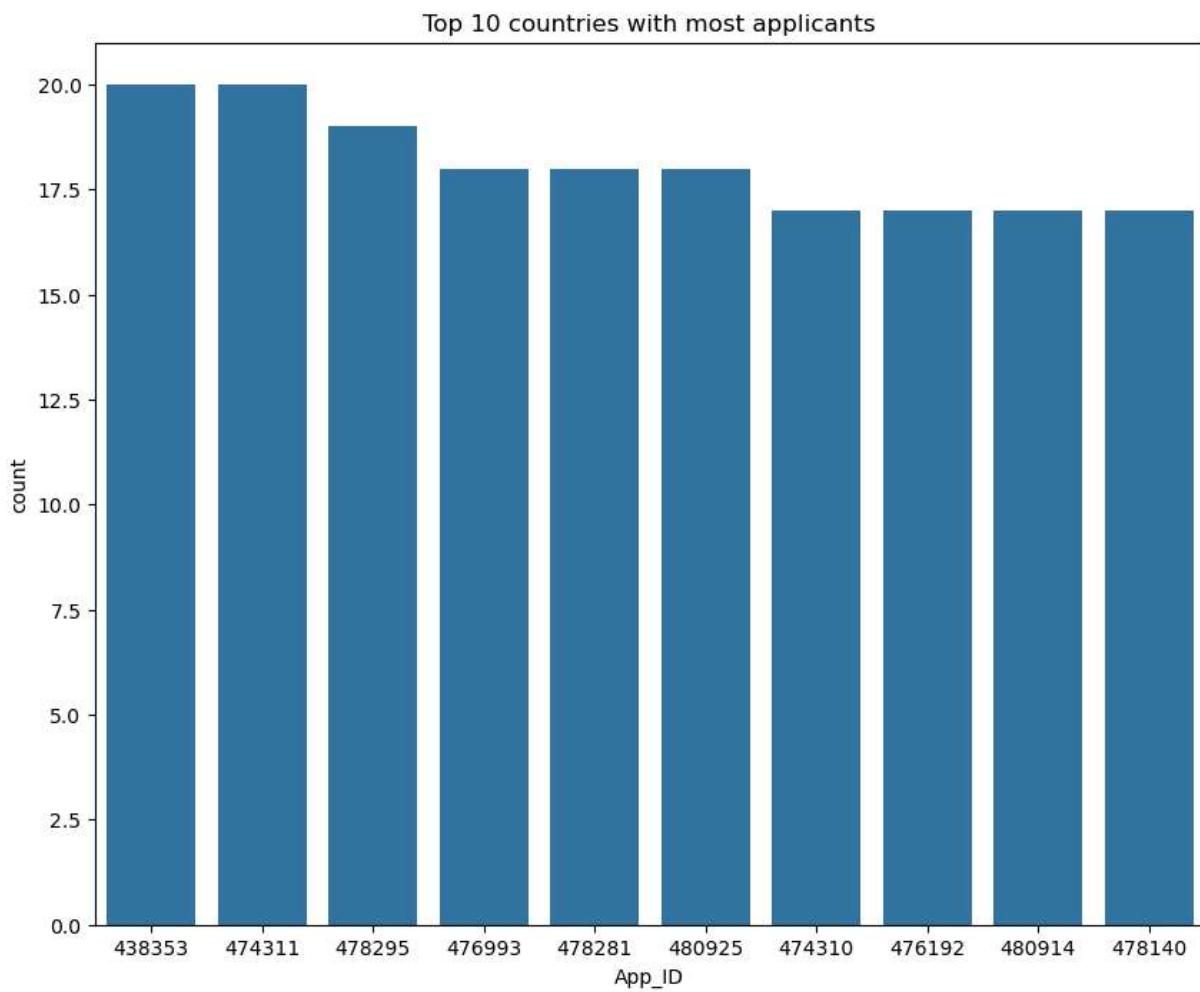
dtype: bool

Plot Graphs

```
In [30]: plt.figure(figsize=(10,5))
sns.countplot(x="Country",data=df,order=df[ "Country"].value_counts().index[:10])
plt.title("Top 10 countries with most applicants")
plt.show()
```



```
In [31]: plt.figure(figsize=(10,8))
sns.countplot(x="App_ID",data=df,order=df[ "App_ID"].value_counts().index[:10])
plt.title("Top 10 countries with most applicants")
plt.show()
```



```
In [32]: !pip install psycopg2-binary sqlalchemy
```

```
Requirement already satisfied: psycopg2-binary in c:\users\satish kumar sharma\conda\envs\gpu-pytorch\lib\site-packages (2.9.10)
Requirement already satisfied: sqlalchemy in c:\users\satish kumar sharma\conda\envs\gpu-pytorch\lib\site-packages (2.0.43)
Requirement already satisfied: greenlet>=1 in c:\users\satish kumar sharma\conda\envs\gpu-pytorch\lib\site-packages (from sqlalchemy) (3.2.4)
Requirement already satisfied: typing-extensions>=4.6.0 in c:\users\satish kumar sharma\conda\envs\gpu-pytorch\lib\site-packages (from sqlalchemy) (4.12.2)
```

```
In [37]: import pandas as pd
from sqlalchemy import create_engine
```

```
# Connection details
user = "postgres"
password = "satish"
host = "localhost"
port = "5400"
database = "postgres"

# Create engine
engine = create_engine(f"postgresql+psycopg2://{user}:{password}@{host}:{port}/{database}")
```

```
In [38]: # Example: read a CSV file
df = pd.read_csv("ApplicantData - ApplicantData.csv")
```

```
# Preview
df.head()
```

Out[38]:

	App_ID	Country	University	Phone_Number
0	12345	India	Illinois Institute of Technology	9823241234
1	12345	India	Illinois Institute of Technology	8805617501
2	12345	India	Illinois Institute of Technology	18019011222
3	347397	Nigeria	Illinois Institute of Technology	7738599513
4	347397	Nigeria	Illinois Institute of Technology	919182706838

In [40]:

```
# Write to SQL (replace 'table_name' with your table name)
df.to_sql("applicant_raw", engine, if_exists="replace", index=False)

print("Data uploaded successfully!")
```

Data uploaded successfully!

In [41]:

```
import pandas as pd
import psycopg2
from io import StringIO

# --- Configuration ---
# Replace with your PostgreSQL credentials
DB_NAME = "applicants_db"
DB_USER = "postgres"
DB_PASS = "your_password" # <-- IMPORTANT: Replace with your password
DB_HOST = "localhost"
DB_PORT = "5432"

# CSV file path
CSV_FILE_PATH = "ApplicantData - ApplicantData.csv"
TABLE_NAME = "applicant_data"

# --- Database Connection and Table Creation ---
try:
    # Establish a connection to the database
    conn = psycopg2.connect(
        dbname=DB_NAME,
        user=DB_USER,
        password=DB_PASS,
        host=DB_HOST,
        port=DB_PORT
    )
    cursor = conn.cursor()

    # Define the SQL command to create the table with the specified schema
    create_table_query = """
CREATE TABLE IF NOT EXISTS {TABLE_NAME} (
    app_id VARCHAR(255),
    country VARCHAR(255),
    """
```

```

        university VARCHAR(255),
        phone_number VARCHAR(255)
    );
"""

# Execute the table creation query
cursor.execute(create_table_query)
conn.commit()
print(f"Table '{TABLE_NAME}' created successfully or already exists.")

# --- Data Loading ---
# Read the CSV file into a pandas DataFrame
df = pd.read_csv(CSV_FILE_PATH, dtype=str).fillna('')

# Use a StringIO object to simulate a file for the COPY command
csv_buffer = StringIO()
df.to_csv(csv_buffer, sep='\t', header=False, index=False)
csv_buffer.seek(0)

# Use psycopg2's copy_from method for efficient bulk insertion
cursor.copy_from(csv_buffer, TABLE_NAME, sep='\t', null='')
conn.commit()
print(f"Data from '{CSV_FILE_PATH}' loaded successfully into '{TABLE_NAME}'.")"

except (Exception, psycopg2.Error) as error:
    print("Error while connecting to or interacting with PostgreSQL:", error)

finally:
    # Close the database connection in the `finally` block to ensure it happens
    if conn:
        cursor.close()
        conn.close()
        print("\nPostgreSQL connection is closed.")

```

Error while connecting to or interacting with PostgreSQL: connection to server at "localhost" (:) failed: Connection refused (0x0000274D/10061)

Is the server running on that host and accepting TCP/IP connections?
connection to server at "localhost" (127.0.0.1), port 5432 failed: Connection refused (0x0000274D/10061)

Is the server running on that host and accepting TCP/IP connections?

```

-----
NameError                                                 Traceback (most recent call last)
Cell In[41], line 63
      59     print("Error while connecting to or interacting with PostgreSQL:", error)
      60
  61 finally:
      62     # Close the database connection in the `finally` block to ensure it happens
---> 63     if conn:
      64         cursor.close()
      65         conn.close()

NameError: name 'conn' is not defined

```

In []: