

```
In [1]: import pandas as pd
```

```
In [2]: df = pd.read_csv('DA-Data-Analyst-interview-assignment-Data.csv')
```

```
In [3]: df.head()
```

Out[3]:

	Uniq Id	Crawl Timestamp	Job Title	Job Salary	Job Experience Required	Key Skills
0	9be62c49a0b7ebe982a4af1edaa7bc5f	2019-07-05 01:46:07 +0000	Digital Media Planner	Not Disclosed by Recruiter	5 - 10 yrs	Media Planning Digital Media
1	3c52d436e39f596b22519da2612f6a56	2019-07-06 08:04:50 +0000	Online Bidding Executive	Not Disclosed by Recruiter	2 - 5 yrs	pre sales closing software knowledge clien...
2	ffad8a2396c60be2bf6d0e2ff47c58d4	2019-08-05 15:50:44 +0000	Trainee Research/ Research Executive- Hi- Tec...	Not Disclosed by Recruiter	0 - 1 yrs	Computer science Fabrication Quality check ...
3	7b921f51b5c2fb862b4a5f7a54c37f75	2019-08-05 15:31:56 +0000	Technical Support	2,00,000 - 4,00,000 PA.	0 - 5 yrs	Technical Support
4	2d8b7d44e138a54d5dc841163138de50	2019-07-05 02:48:29 +0000	Software Test Engineer - hyderabad	Not Disclosed by Recruiter	2 - 5 yrs	manual testing test engineering test cases ...

```
In [4]: df.columns.values.tolist()
```

```
Out[4]: ['Uniq Id',  
         'Crawl Timestamp',  
         'Job Title',  
         'Job Salary',  
         'Job Experience Required',  
         'Key Skills',  
         'Role Category',  
         'Location',  
         'Functional Area',  
         'Industry',  
         'Role']
```

```
In [5]: df.columns.values.tolist()
```

```
Out[5]: ['Uniq Id',  
         'Crawl Timestamp',  
         'Job Title',  
         'Job Salary',  
         'Job Experience Required',  
         'Key Skills',  
         'Role Category',  
         'Location',  
         'Functional Area',
```

```
'Industry',  
'Role']
```

What are the Job Opportunities in the Top 20 cities?

```
In [6]: df['Job Title'].value_counts()[:20]
```

```
Out[6]: Business Development Executive    120  
Sales Executive                          120  
PHP Developer                           117  
Business Development Manager            99  
Graphic Designer                        84  
Software Engineer                       79  
Java Developer                          77  
Android Developer                       77  
Content Writer                          74  
Sales Manager                           72  
Business Analyst                        62  
Senior Software Engineer                61  
Marketing Executive                     59  
Web Designer                            58  
Dot Net Developer                       57  
Job Description                         57  
Software Developer                      46  
Project Manager                         43  
HR Executive                            42  
Accounts Executive                      40  
Name: Job Title, dtype: int64
```

```
In [7]: df['Location'].value_counts()[:20]
```

```
Out[7]: Bengaluru          4986  
Mumbai                    3318  
Pune                      2431  
Hyderabad                 2144  
Gurgaon                   1715  
Chennai                   1564  
Kolkata                   1339  
Delhi                     1212  
Noida                     1141  
Ahmedabad                 1032  
Delhi NCR                  734  
Chandigarh                 304  
Bengaluru                  229  
Navi Mumbai                167  
Delhi NCR,Gurgaon          141  
Hyderabad                  137  
Faridabad                  121  
Delhi NCR,Gurgaon,Noida    113  
Mumbai,Mumbai Suburbs      93  
Ghaziabad                  84  
Name: Location, dtype: int64
```

```
In [8]: strip = lambda x: x.strip()
```

```
top_20_city_data = df[df['Location'].isin(df['Location'].value_counts().keys()).tolis  
top_20_city_data['Location'] = top_20_city_data['Location'].apply(strip)
```

<ipython-input-8-15883b6dcd1a>:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: <https://pandas.pydata.org/pandas-docs/stable/u>

```
ser_guide/indexing.html#returning-a-view-versus-a-copy
top_20_city_data['Location'] = top_20_city_data['Location'].apply(strip)
```

```
In [9]: df['Location'].value_counts()[:20]
```

```
Out[9]: Bengaluru          4986
Mumbai          3318
Pune            2431
Hyderabad       2144
Gurgaon         1715
Chennai         1564
Kolkata         1339
Delhi           1212
Noida           1141
Ahmedabad       1032
Delhi NCR        734
Chandigarh       304
Bengaluru        229
Navi Mumbai     167
Delhi NCR,Gurgaon 141
Hyderabad        137
Faridabad        121
Delhi NCR,Gurgaon,Noida 113
Mumbai,Mumbai Suburbs 93
Ghaziabad        84
Name: Location, dtype: int64
```

```
In [10]: from mpl_toolkits import basemap
print(basemap.__version__)
```

1.3.3

```
In [11]: import matplotlib.pyplot as plt
from mpl_toolkits.basemap import Basemap
import numpy as np
```

```
In [12]: fig = plt.figure(figsize=(12,6))
m = Basemap(
    projection='mill',
    resolution='c',
    llcrnrlat=7.890072140756692,
    llcrnrlon=64.6128442670092,
    urcrnrlat=39.31741853835297,
    urcrnrlon=97.64121300133036
)
m.drawcoastlines() # to draw continental coast lines
m.drawcountries(color='r')
m.drawmapboundary(color='orange',linewidth=5)
top_20_city = ['Bengalore', 'Mumbai', 'Pune', 'Hyderabad', 'Gurgaon', 'Chennai', 'Kolkata']
longitudes = [77.580643, 72.877426, 73.856255, 78.491684, 77.026344, 80.237617, 88.363892,
latitudes = [12.972442, 19.076090, 18.516726, 17.387140, 28.457523, 13.067439, 22.572645, 2
x,y = m(longitudes,latitudes)
plt.scatter(x,y,s=15,color='red')
for i,label in enumerate(top_20_city):
    plt.text(x[i]+100000,y[i],label)
plt.show()
```



What are the Job Opportunities in the Top 20 Industries?

```
In [13]: df['Job Title'].value_counts()[:20]
```

```
Out[13]: Business Development Executive    120
Sales Executive                          120
PHP Developer                           117
Business Development Manager            99
Graphic Designer                        84
Software Engineer                       79
Java Developer                          77
Android Developer                       77
Content Writer                          74
Sales Manager                           72
Business Analyst                         62
Senior Software Engineer                 61
Marketing Executive                      59
Web Designer                            58
Dot Net Developer                       57
Job Description                          57
Software Developer                       46
Project Manager                         43
HR Executive                            42
Accounts Executive                       40
Name: Job Title, dtype: int64
```

```
In [14]: df['Location'].value_counts()[:20]
```

```
Out[14]: Bengaluru    4986
Mumbai               3318
Pune                 2431
Hyderabad            2144
Gurgaon              1715
Chennai              1564
Kolkata              1339
Delhi                1212
Noida                1141
Ahmedabad            1032
Delhi NCR             734
Chandigarh           304
Bengaluru            229
```

```

Navi Mumbai          167
Delhi NCR,Gurgaon    141
  Hyderabad          137
Faridabad             121
Delhi NCR,Gurgaon,Noida 113
Mumbai,Mumbai Suburbs 93
Ghaziabad             84
Name: Location, dtype: int64

```

```

In [15]: strip = lambda x: x.strip()

top_20_industry_data = df[df['Industry'].isin(df['Industry'].value_counts().keys()).t

```

```

In [16]: top_20_industry_data['Industry'].value_counts()[:20]

```

```

Out[16]: IT-Software, Software Services          10012
Recruitment, Staffing                          2721
BPO, Call Centre, ITes                         1754
Banking, Financial Services, Broking           1685
Education, Teaching, Training                 1364
IT-Software / Software Services                786
Medical, Healthcare, Hospitals                 676
Internet, Ecommerce                           585
Construction, Engineering, Cement, Metals      566
Strategy, Management Consulting Firms           532
Travel , Hotels , Restaurants , Airlines , Railways 510
Automobile, Auto Anciliary, Auto Components    498
Pharma, Biotech, Clinical Research             466
Media, Entertainment, Internet                 459
Industrial Products, Heavy Machinery            459
Other                                           424
Advertising, PR, MR, Event Management          403
Real Estate, Property                         393
Insurance                                      343
Oil and Gas, Energy, Power, Infrastructure      317
Name: Industry, dtype: int64

```

```

In [ ]:

```

What are the Top 20 Job Categories

```

In [17]: df['Role Category'].value_counts()[:20]

```

```

Out[17]: Programming & Design          8551
Voice                                  2385
Retail Sales                          2359
HR/ Recruitment / IR                  1133
Admin/Maintenance/Security/Datawarehousing 928
Accounts                             882
Senior Management                     880
Other                                 865
Corporate Sales                       704
Production/Manufacturing/Maintenance  592
System Design/Implementation/ERP/CRM  507
Marketing                             487
Medical Professional                  438
Creative                              396
Engineering Design                    368
Finance/Audit                         339
Teachers                             329
Project Management                    311
University Level                      305
Back Office/Web/Transaction Processing 294
Name: Role Category, dtype: int64

```

What are the Top 20 Job Roles?

```
In [18]: df['Role'].value_counts()[:20]
```

```
Out[18]: Software Developer          6396
Associate/Senior Associate -(NonTechnical)  1728
Sales/Business Development Manager  1561
Sales Executive/Officer            1168
Other                             896
Team Lead/Technical Lead          847
Associate/Senior Associate -(Technical)  513
Accounts Executive/Accountant      511
Recruitment Executive              427
Testing Engineer                   408
Fresher                           380
Technical Architect                366
System Administrator               318
HR Executive                       297
Business Analyst                   268
Design Engineer                   230
Telecalling/Telemarketing Executive  227
Head/VP/GM/National Manager -Sales  216
Project Manager-IT/Software        215
Marketing Manager                  207
Name: Role, dtype: int64
```

What is the Top 5 Location for Top 20 Industries?

```
In [19]: top_20_industry_data['Location'].value_counts()[:5]
```

```
Out[19]: Bengaluru    4332
Mumbai              2729
Pune                2155
Hyderabad           1858
Gurgaon             1453
Name: Location, dtype: int64
```

What are the Skills required for various job categories?

```
In [20]: strip = lambda x: str(x).strip()
col_preprocess = ['Key Skills', 'Role Category']

for col in col_preprocess:
    df[col] = df[col].apply(strip)
```

```
In [21]: df.groupby(['Role Category'])['Key Skills'].transform(lambda x: ','.join(x)).tolist()
```

```
Out[21]: 'Media Planning| Digital Media,nan,Internship| Digital marketing| Social media| Relationship building| Animation| Creative| Indesign| Business Executive| Presentation skills| Senior,event production| Client Servicing| event operation| mice| events| conferences| event coordinator,Excel| Financial markets| corporate events| Public speaking| Digital marketing| Powerpoint| Marketing communication| Marketing Executive| Event Executive| Marketing,Campaigns| Digital Marketing| Project Planning,Account Manager| Client Servicing| Key account manager| Account Management| Digital Strategy| brand solutions| brand management,Client Servicing| relationship management| business development,Market research| Digital marketing| Social media| SEM| Print media| Analytics| Monitoring| Display advertising| UX| Analytical skills,customer service| customer support| ms office| client servicing| relationship management| Client Relationship Management| client relation| customer relation| customer care| cse,Marketing| Ev
```

ents| Sales| Presentation Skills| Exhibitions| Fund Raising| Communication Skills,digital marketing|online marketing|sem|search engine marketing|pay per click|google ad words professional|media planning|social media marketing|programmatic|paid media|digital campaigns|campaign management|campaign planning|Google Analytics|DCB,Social Media Executive| Digital Marketing Executive| facebook marketing| Management,Mba Fresher,Client Servicing| Project Manager| Project Manager Event| Client s Executive| Manager Client Servicing| Client Associate| Event Manager| Assistant Event Manager| Brand Activation| Activation Officer,Relationship management| Brand communication| Management,Social media marketing| Customer support,Social media marketing| SEO| Digital marketing| Email marketing| Team management| Google Analytics| Monitoring| Campaign planning| marketing automation| digital strategy,Internet Advertising| Online Promotion| Freelancer| Online Marketing| Digital Marketing| Fresher| Part Time,FMCG Marketing| Promotions| Events| Communications skills| Brand Promotion| Marketing Executive| Presentation skills| Ms office,Social media marketing| Display advertising| Web technologies| SMO| Analytical| Marketing campaigns| SEO| Digital marketing| Search engine optimization,seo| social media| marketing strategy| social media marketing| facebook| twitter| analytics| social marketing| inbound| qualitative,English| Social marketing| Bidding| Email| Email marketing| Campaign management| Analytical| Service| Advertising| Data entry,Business Executive,Advertising,social media marketing| seo| campaigns| digital advertising| sem| website management| facebook| promotions| twitter| campaign management| brand| atl| Marcom| branding| digital marketing| online marketing,Credit Officers| Risk Management| Administration| Report Generation| Policies| Analysts| analytical skills,nan,Sales| Cold Calling| Business Development,Business Analyst| English| Tracking| CV| Manager Client Servicing| Writing| Relationship| Agency| mass| Hons,Social media marketing| LinkedIn| Content strategy| CV| Social media| consumer marketing| Campaigns| Digital marketing| Facebook| Business Executive,accounts executive| Brand Manager| client servicing| client management| content strategy| accounts manager,Tracking| Online media| Negotiation skills| Media planning| Negotiation| Support| Compliance| Budgeting| Writing| Agency,btl activation| Atl| marketing management| events| Marketing,Social media marketing| Web technologies| SMO| Web analytics| Networking| Corporate| content marketing| Management| Brand building| Monitoring,nan,sales account| digital marketing| social media| seo| client servicing| strategy| account management| digital strategy| presales,online promotion| seo executive| web development| work from| internet marketing| freelancer| online advertising| part time,Market analysis| Backend| Client satisfaction| Brand activation| Tracking| Conceptualization| Strategic business planning| Cross selling| Business Executive| Project execution,Social Media Manager| Social Media Marketing| Social Media Optimization,marketing| international business| Business Development| client servicing| International Marketing| international business development| Marketing Manager| Marketing Management| business development management,Digital Marketing| HR| Typing| Computer Operating| Human Resource Management| Hospitality| Facebook Marketing| Telecalling| telemarketing,English| advertising agency| SMM| Content| Writing| SEM| SEO| Business Executive,Media| Billing| Email| closure| Tracking| Internal communication| Reconciliation| Cross selling| confirmation| Client Service Executive,Social media| Budgeting,Public relations| Typing| Product service| Marketing campaigns| Distribution| sponsorship| Time| Research| Marketing Executive| Advertising,nan,nan,Profit head| Curriculum| Journalism| Trainee| Database| Research| Education| Publishing| Entry Level| Technology,Market research| Monitoring| Strategic thinking| National| Support| Report preparation| Target| Trainee| Regional Coordinator,Social media marketing| Analytical skills| Lead generation| Usage| Management,operations| sound| setup| handling| btl activities| travel| events,Sales Coordinator| Marketing Consultant| Sales Support| Sales Advisor,Media research| Agency sales| Revenue generation| Radio| Station| Printing,Sales| Telecalling| Back office| Marketing Coordinator| Vendor| Counselling,nan,analytics| algorithms| ppc| smm| adwords| smo| team lead| sem| company| seo,BPO| Social media marketing| Digital media| Consulting| HTML| Customer relationship| ITES| Management| CRM,PPC| Copy writing| Internet marketing| SEM| Pay per click| Management| Merchandising| Facebook| international clients| Analytics,Sales| Social Media| Proof Reading| Editing| Team Leading| Campaigns| Copy Writing,Analytics| HTML| SEM| Javascript| PPC| Event management| Digital marketing| Customer acquisition| Tools| Algorithms,Trade| Event Executive,Digital media| Email marketing| Digital design| Social media| Production operations| Billing| CMS| digital strategy| SEO| Digital marketing,Key Accounts Management| business growth| key accounts,Team Leader,Media Management| LinkedIn| Social Media| blogs| Youtube| Twitter| Facebook| content,Trade| Online media| Usage| Media strategy| Social media| Brand awareness| Research| Management| Business Executive| Analytics,Video Editing| Video,Digital Marketing| Advertising| Design| Media Buying| Loyalty Programs| Brand Strategy| Mystery Shopping| 3D Rendering| DTP Operating| Real Estate,social media| scheduling| facebook| twitter| coordinating| planning,Event Management| Marketing Collaterals| Marketing,Display advertising| Google Analytics| Social media| Javascript| Google AdWords| SEM| HTML| Omniture| SEO

| Digital marketing,digital marketing| client relationship management| new| managem
nt skills| team management skills| crisis management| events| quality,Online marketi
ng| SEO| Analytical| Digital marketing| Social media| Analytics| HTML| Javascript| D
isplay advertising| Marketing campaigns,Media Buying| Media Planning| Advertising,Le
ad Generation| Customer Handling| Post Sales| Marketing| Material Receipt,marketing
executive| distributors| customer profiling| Business Development| sales executive,m
arket research| customer profiling| sales executive| marketing executive| distributo
rs| Business Development,Digital Marketing Executive| Strategy making| Sound| Creati
ve| Compensation| Usage,Executive| Excel| Research| Management| Fact| Media planning
| Reporting tools| Monitoring,Social media| PCM| Marketing| Digital Marketing,softwa
re development| online promotion| Fresher| work from| freelancer| online advertising
| part time,bpo| digital marketing| online marketing| fresher| online branding| web
marketing| business development| kpo| seo| internet marketing| freelancer| part tim
e,customer marketing| Excel| Inventory management| UPS| MS Office| Powerpoint| Busin
ess Executive| Sales support,Manager Client Servicing| Client Service Executive| Sil
icon| Advertising,digital marketing| team handling,Branded Content| Video Editing| A
dobe Premiere| Post Production| Motion Graphics| Social Media| Science| Final Cut Pro
| Color Correction| Script Writing,Digital media| Integrated marketing| Client mana
gement| Analytics| Media planning| Marketing| Investment,PPC| Display advertising| E
mail marketing| SMO| Online marketing| Internet marketing| SEM| SEO| Digital marketi
ng,nan,Advertising| Media Relations| Internal Communication| Branding| Customer Supp
ort| Customer Service| Twitter| Facebook| Social Media| Customer Satisfaction| Linke
din| Content Writing,online promotion| seo executive| web development| work from| in
ternet marketing| freelancer| online advertising| part time,open source| website| di
gital media| charts| drafting| gap analysis| documentation| english| sdlc| process f
low,Entry Level| Publishing| Trainee| Education| Technology| Database| Journalism| R
esearch,assembly language| Six Sigma| JavaScript| .NET| HTML| Oracle| Python| SQL,Pr
esales| Customer Satisfaction| Excel| Powerpoint,Ad Sales| B2B Sales| Corporate Sale
s,SEM| Google AdWords| Web analytics| Analytical|digital marketing head| Search eng
ine marketing| SEM Executive| Advertising| Research| Business Executive,blog writing
| copy writing| blogs| service marketing| content writing| seo writing| product serv
ice| case studies| articles| social media,Social media| Email| LinkedIn| Digital Mar
keting Executive| Business Analyst| Project Coordinator| Client interaction| Spoken
english| Technical| Business Executive,Client Onboarding| Customer Experience| Campa
ign Management| Marketing| Customer Engagement,Customer Relationship| Customer Care|
Relationship Management| Receptionist Activities,Digital Advertising| Sales,nan,seo|
digital marketing| dealing| email marketing| sem| social media marketing| digital me
dia| affiliate marketing| google adwords| google analytics,operations| customer sati
sfaction| leadership skills| complaints| microsoft excel| quality,Social media| Publ
ishing| Email| Social media marketing| iPhone| Facebook| Photoshop| Tracking| Google
Analytics| Content Editor,nan,development| operations| relationship management| stra
tegy| brand ing| coordinating| conceptualization| ideation| team,Corporate Gifts| Pa
ckaging| packaging development,digital marketing| Social Work| computer operating| o
nline marketing| office assistant| social media marketing| back office executive| bu
siness development| human resource management| online advertising,Branding| SEO| Pub
lishing| Social media| SEM| Web designing| HTML| Photoshop| Facebook,brand manager|
account manager| client servicing| advertising| account management,Social media mark
eting| Marketing campaigns| Management| Marketing Executive,Social media marketing,S
ocial Media Executive| Interpersonal skills| Typing speed| E-commerce| Troubleshooti
ng| MIN,events| activation| event operations| event manager| vendor management| Even
t Management| Operations Manager| Btl Activation| Btl,BPO| Online Branding| customer
relationship| Digital Marketing| kpo| internet marketing| part time| Freelancer| Onl
ine Marketing| Fresher| Web Marketing| Business Development| seo,Marketing Strategy|
Twitter| Facebook| Brand Campaigns| Social Media| Customer Engagement| Buyer Activit
ies| Customer Service,Online Promotion| BPO| online marketing| Freelancer| Internet
Marketing| MBA Fresher| web marketing| Business Development| Part Time,Manager Quali
ty Assurance| Sales process| Event Coordinator| Supervisor| Packaging| Room| Managem
ent,Copy Writing| Content Writing| Proof Reading| Social Media| SEO| Digital Marketi
ng| English Language| Customer Care| Content| content writer| content developer| con
tent editor| creative writer| creative content,Business operations| design studio| A
dvertising| Executive,photoshop| corel draw| 2d graphics| 3d graphics,Team Managemen
t| Channel Partners| Customer Relationship| Project Sales| Client Relationship Manag
ement| New Business Development| Strategy Formulation| Sales Management| Marketing|
Client Handling,SEO| Social media| Content writing| Analytical skills| Customer mana
gement| Head Business Development| Management| Basic| Relationship,marketing| bdm| c
lient servicing| client interaction| sales coordination| back office| fresher| opera
tion| excel| mis| email| marketing executive| client onboarding,digital marketing| m
arketing strategy| marketing communication| marketing analysis| client servicing| ma
rketing initiatives,market research| customer profiling| sales executive| marketing

What is the Average salary offered in the top 10 cities?

```
In [22]: top_20_city_data = df[df['Location'].isin(df['Location'].value_counts().keys().tolist())]
top_20_city_data['Location'] = top_20_city_data['Location'].apply(strip)
```

```
<ipython-input-22-f84275e8d890>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
top_20_city_data['Location'] = top_20_city_data['Location'].apply(strip)
```

```
In [23]: top_20_city_data['Location'].value_counts()[:10].mean()
```

```
Out[23]: 2088.2
```

What is the Average salary offered by the top 10 Industries in the top 5 locations?

```
In [24]: top_20_city_data['Industry'].value_counts()[:10].mean()
```

```
Out[24]: 1535.4
```