

Final Project

Ahona Roy

2001-12-01

```
RendererRegistry.enable('png')
```

	ID	Experience_Years	Age	Gender	Salary
0	1	5	28	Female	250000
1	2	1	21	Male	50000
2	3	3	23	Female	170000
3	4	2	22	Male	25000
4	5	1	17	Male	10000

	JobTitle	Gender	Age	PerfEval	Education	Dept \
0	Graphic Designer	Female	18	5	College	Operations
1	Software Engineer	Male	21	5	College	Management
2	Warehouse Associate	Female	19	4	PhD	Administration
3	Software Engineer	Male	20	5	Masters	Sales
4	Graphic Designer	Male	26	5	Masters	Engineering

	Seniority	BasePay	Bonus
0	2	42363	9938
1	5	108476	11128
2	5	90208	9268
3	4	108080	10154
4	5	99464	9319

Testing

Hello, World!

Delete this

Data cleaning

```
JobTitle      0
Gender        0
Age           0
PerfEval      0
Education     0
Dept          0
Seniority     0
BasePay       0
Bonus         0
dtype: int64
```

```
ID            0
Experience_Years 0
Age           0
Gender        0
Salary        0
dtype: int64
```

		Age	PerfEval	Seniority	BasePay	Bonus
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	41.393000	3.037000	2.971000	94472.653000	6467.161000	
std	14.294856	1.423959	1.395029	25337.493272	2004.377365	
min	18.000000	1.000000	1.000000	34208.000000	1703.000000	
25%	29.000000	2.000000	2.000000	76850.250000	4849.500000	
50%	41.000000	3.000000	3.000000	93327.500000	6507.000000	
75%	54.250000	4.000000	4.000000	111558.000000	8026.000000	
max	65.000000	5.000000	5.000000	179726.000000	11293.000000	

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype
---  -
0   JobTitle    1000 non-null   object
1   Gender      1000 non-null   object
2   Age         1000 non-null   int64
```

```

3   PerfEval    1000 non-null    int64
4   Education   1000 non-null    object
5   Dept        1000 non-null    object
6   Seniority   1000 non-null    int64
7   BasePay     1000 non-null    int64
8   Bonus       1000 non-null    int64
dtypes: int64(5), object(4)
memory usage: 70.4+ KB

```

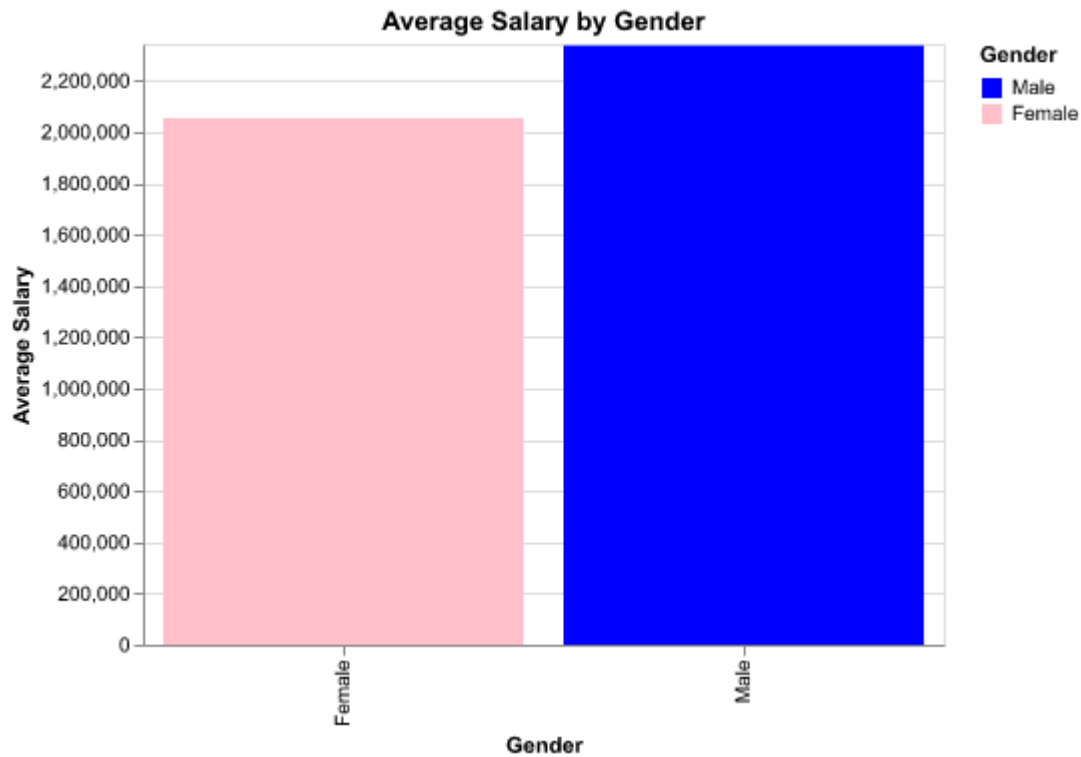
```

#Renaming columns
#Creating total compensation column
#Sorting based on seniority level
#Clean up dataset gender_3

```

Data visualization

1.Bar Chart: Average Salary by Gender



2.Scatter Plot: Experience vs. Salary



3.Geopandas Map Visualization

	intnum68	pernum68	wave	sex	intnum	farminc	region	famwgt	relhead	\
0	1152	24	1981	1	3505	0.0	1.0	17.000	10	
1	5057	171	1981	1	675	0.0	2.0	2.000	10	
2	6375	170	2009	1	8505	0.0	1.0	6.565	10	
3	2261	3	1999	1	4202	0.0	2.0	25.973	10	
4	5927	7	1990	1	605	0.0	3.0	14.819	10	

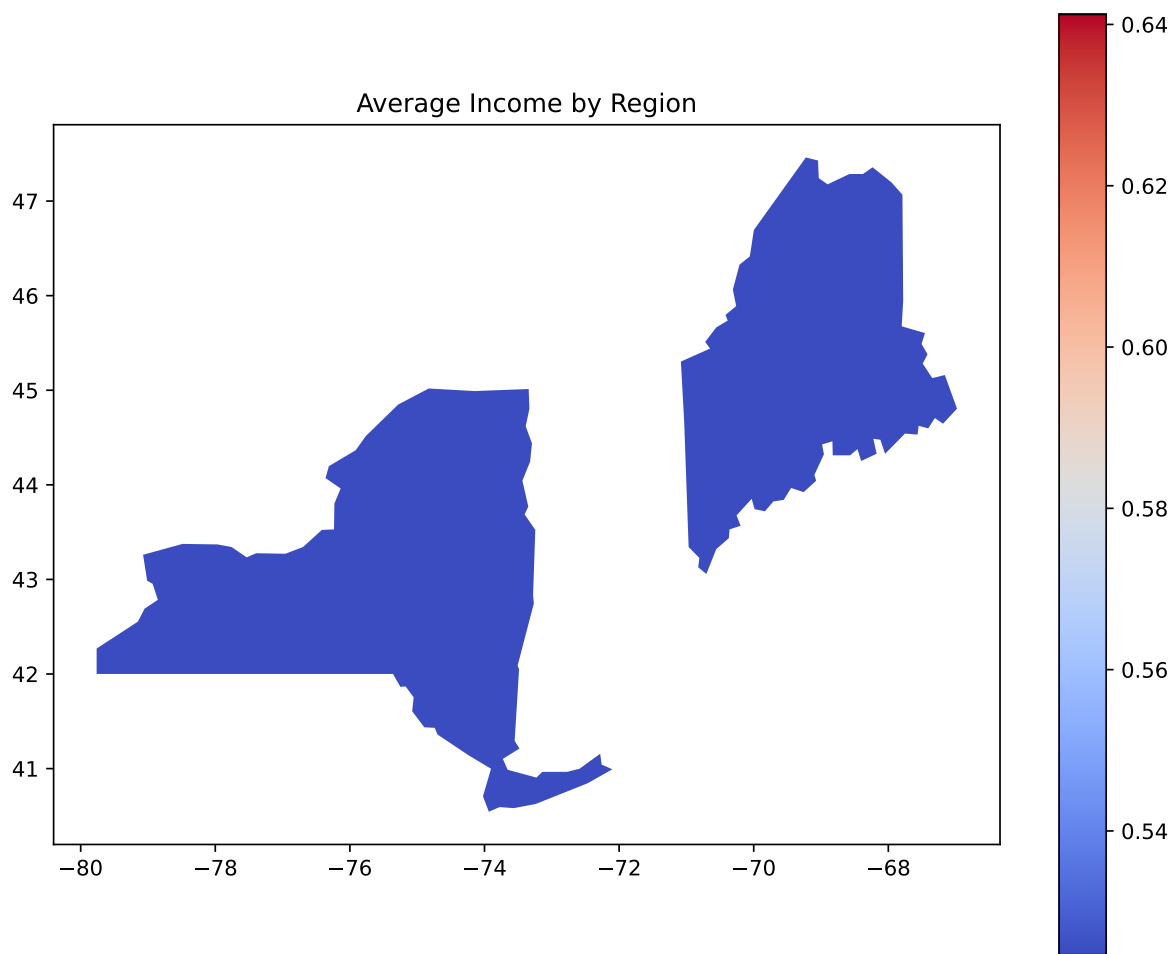
	age	...	foodcare	building	sales	officeadmin	farmer	\
0	34	...	0	0	1	0	0	
1	32	...	0	0	0	0	0	
2	64	...	0	0	0	0	0	
3	50	...	0	0	0	0	0	
4	26	...	0	0	0	0	0	

	constructextractinstall	production	transport	sumocc	LEHS
0	0	0	0	1	0
1	0	1	0	1	1
2	0	1	0	1	1
3	0	0	1	1	1
4	0	1	0	1	1

```
[5 rows x 274 columns]
```

```
Index(['intnum68', 'pernum68', 'wave', 'sex', 'intnum', 'farminc', 'region',  
      'famwgt', 'relhead', 'age',  
      ...  
      'foodcare', 'building', 'sales', 'officeadmin', 'farmer',  
      'constructextractinstall', 'production', 'transport', 'sumocc', 'LEHS'],  
      dtype='object', length=274)
```

```
[ 1.  2.  3.  4. nan]
```



```
intnum68  pernum68  wave  sex  intnum  farminc  region  famwgt  relhead  \
```

0	1152	24	1981	1	3505	0.0	1.0	17.000	10
1	5057	171	1981	1	675	0.0	2.0	2.000	10
2	6375	170	2009	1	8505	0.0	1.0	6.565	10
3	2261	3	1999	1	4202	0.0	2.0	25.973	10
4	5927	7	1990	1	605	0.0	3.0	14.819	10

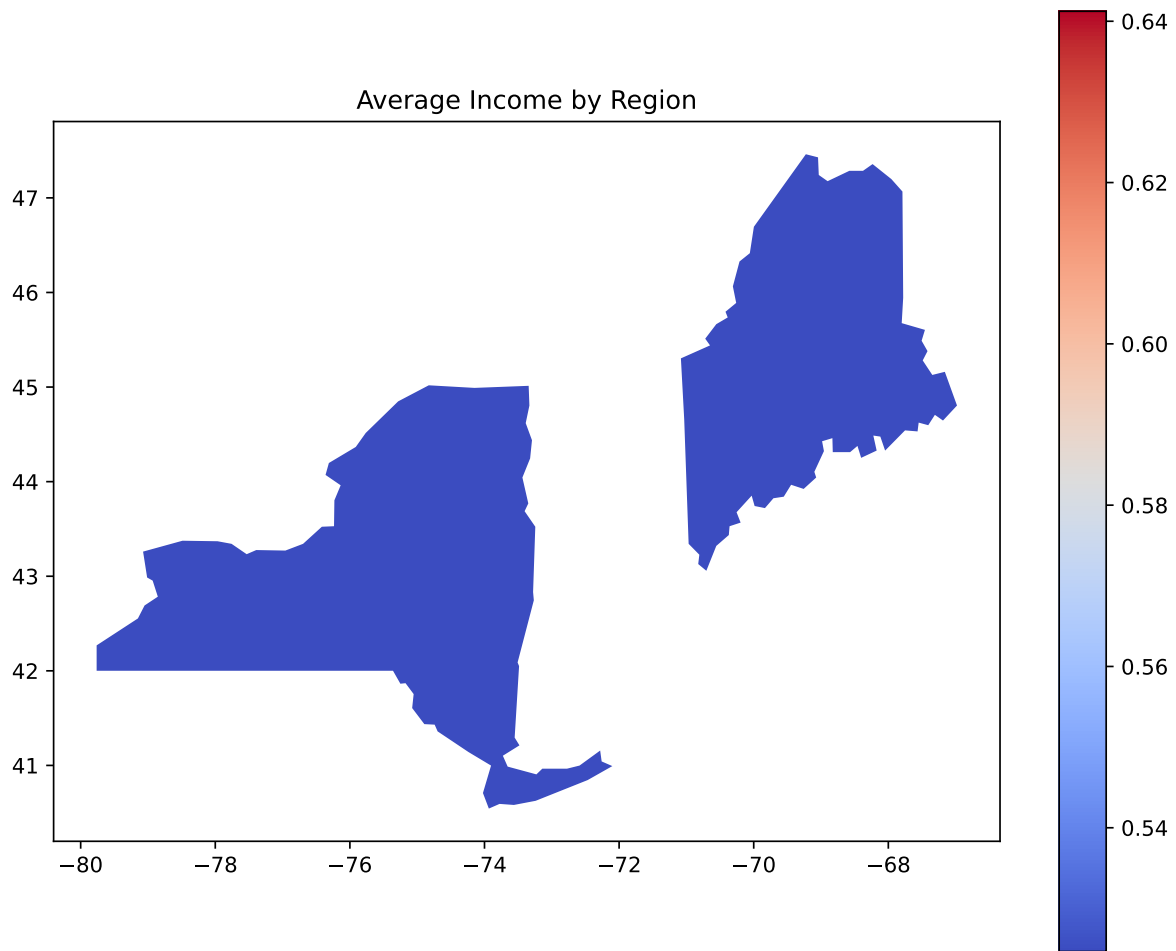
	age	...	foodcare	building	sales	officeadmin	farmer	\
0	34	...	0	0	1	0	0	
1	32	...	0	0	0	0	0	
2	64	...	0	0	0	0	0	
3	50	...	0	0	0	0	0	
4	26	...	0	0	0	0	0	

	constructextractinstall	production	transport	sumocc	LEHS
0	0	0	0	1	0
1	0	1	0	1	1
2	0	1	0	1	1
3	0	0	1	1	1
4	0	1	0	1	1

[5 rows x 274 columns]

```
Index(['intnum68', 'pernum68', 'wave', 'sex', 'intnum', 'farminc', 'region',
      'famwgt', 'relhead', 'age',
      ...,
      'foodcare', 'building', 'sales', 'officeadmin', 'farmer',
      'constructextractinstall', 'production', 'transport', 'sumocc', 'LEHS'],
      dtype='object', length=274)
```

[1. 2. 3. 4. nan]



	id	name	density	geometry \
0	01	Alabama	94.650	POLYGON ((-87.3593 35.00118, -85.60668 34.9847...
1	02	Alaska	1.264	MULTIPOLYGON (((-131.60202 55.11798, -131.5691...
2	04	Arizona	57.050	POLYGON ((-109.0425 37.00026, -109.04798 31.33...
3	05	Arkansas	56.430	POLYGON ((-94.47384 36.50186, -90.15254 36.496...
4	06	California	241.700	POLYGON ((-123.23326 42.00619, -122.37885 42.0...

	region_name	avg_income
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN

	region_name	avg_income
--	-------------	------------

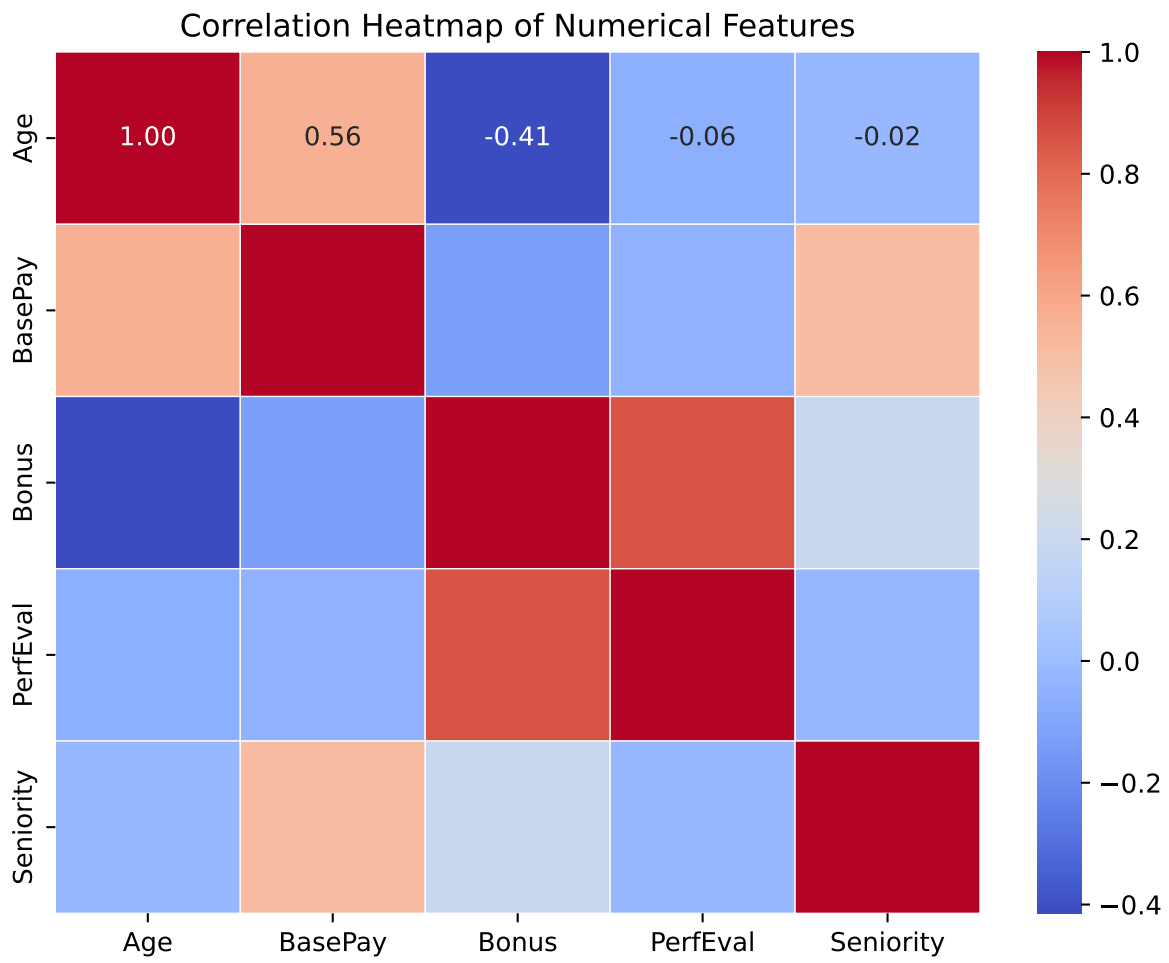
19	Northeast	0.582945
32	Northeast	0.582945

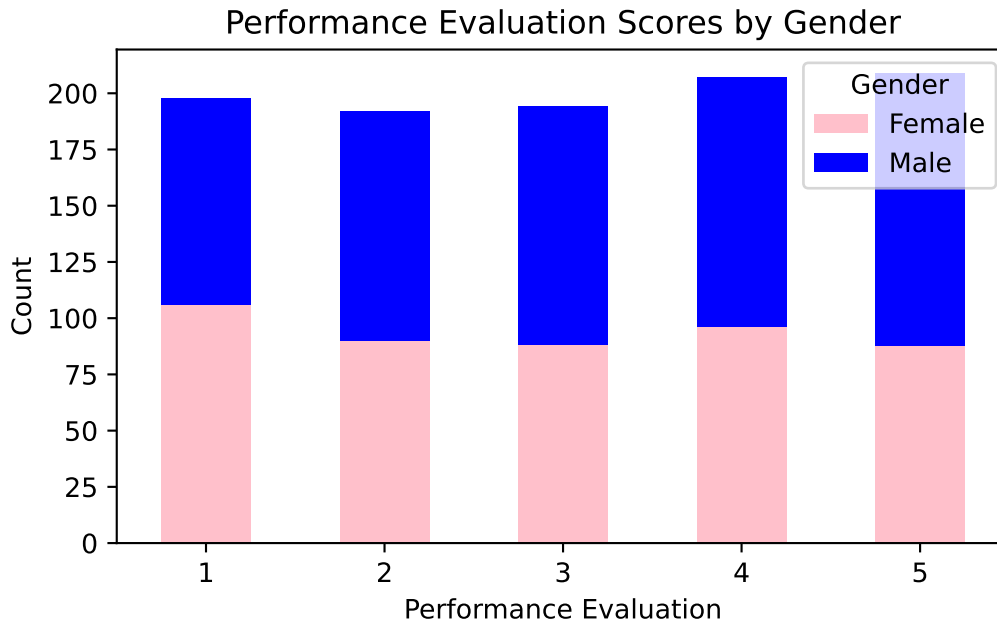
Dataset 1 data visualization

	JobTitle	Gender	Age	PerfEval	Education	Dept \
0	Graphic Designer	Female	18	5	College	Operations
1	Software Engineer	Male	21	5	College	Management
2	Warehouse Associate	Female	19	4	PhD	Administration
3	Software Engineer	Male	20	5	Masters	Sales
4	Graphic Designer	Male	26	5	Masters	Engineering

	Seniority	BasePay	Bonus
0	2	42363	9938
1	5	108476	11128
2	5	90208	9268
3	4	108080	10154
4	5	99464	9319

Heatmap





Extra Credit

`['we', 'are', 'looking', 'for', 'a', 'dynamic', 'leader', 'to', 'manage', 'our', 'sales', 'to']`

```
[nltk_data] Downloading package punkt to /Users/ahonaroy/nltk_data...
[nltk_data] Package punkt is already up-to-date!
```

#1.

```
[nltk_data] Downloading package punkt to /Users/ahonaroy/nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data] /Users/ahonaroy/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data] /Users/ahonaroy/nltk_data...
[nltk_data] Package wordnet is already up-to-date!
```

```
Male words: []
Female words: []
```

#2.

['we', 'are', 'looking', 'for', 'hire', 'experts', 'flutter', 'developer', '.', 'so', 'you',
['python/django', '(', 'developer/lead', ')', '-', 'job', 'code', '(', 'pdj', '-', '04', ')',
['data', 'scientist', '(', 'contractor', ')', 'bangalore', ',', 'in', 'responsibilities', 'w',
, 'all', 'the', 'way', 'through', 'production', '.', 'qualifications', 'at', 'least', '3',
, 'advantage', 'experience', 'with', 'big', 'data', '-', 'advantage', 'experience', 'with',
, 'big', 'advantage', 'experience', 'with', 'deep', 'learning', 'frameworks', '(', 'pytorch',
, 'advantage', '.', 'experience', 'in', 'the', 'telecommunication', 'domain', 'and/or', 'fr',
['job', 'description', ':', 'strong', 'framework', 'outside', 'of', 'ios', 'is', 'always', 'a',
['job', 'responsibility', 'full', 'stack', 'engineer', '-', 'react', 'role', 'make', 'impact',
for', 'love', 'pet', 'click', 'meet', 'mike', 'goodwin', 'petsmart', '', 'cio', '', 'love',
productivity', 'finest', '!', '"', 'lil', '', 'paw', '"', 'learning', 'center', 'onsite', 'c',
['software', 'developer', '-', 'integration', '*', 'immediate', 'opening', '!', '*', 'a', 'd',
['senior', 'full', 'stack', 'developer', '\\\\-', '1800026h', 'cwt', 'looking', 'senior', 'full',
['job', 'description', ':', 'reactjs', '+', 'nodejs', ',', 'azure', 'functions', ',', 'and',
['main', 'responsibilities', 'and', 'deliverables', ':', 'manage/support', 'the', 'rollout',
['overview', 'based', 'in', 'silicon', 'valley', ',', 'tintri', 'is', 'a', 'wholly', 'owned',
, 'rest', 'api', 'dev', 'job', 'description', 'responsibilities', ':', 'we', 'are', 'looking',
['richardsontexasunited', 'state', 'bausp4', 'boeing', 'company', 'currently', 'seeking', 'd',
['remote', ',', 'any', 'where', '2', '-', '4', 'years', '8l', '-', '12l', '/', 'yr', '(', 'o',
, 'opportunity', 'to', 'lead', 'development', 'of', 'a', 'global', 'tech', 'product', ',',
['excited', 'by', 'big', 'data', ',', 'machine', 'learning', 'and', 'predictive', 'software',
['as', 'a', 'member', 'of', 'our', 'software', 'engineering', 'group', 'you', 'will', 'dive',
['job', 'description', 'job', 'title', ':', 'software', 'engineer', 'i', 'summary', 'the', 's',
['overview', ':', 'western', 'national', 'is', 'seeking', 'a', 'highly', 'collaborative', 'an',
, '2016', ')', ',', 'and', 'previously', 'as', 'a', 'recipient', 'of', 'the', 'minnesota',
, 'we', 'offer', 'affordable', '(', 'thanks', 'to', 'a', 'strong', 'employer', 'subsidy', 'x',
, 'including', 'on-site', 'massage', ',', 'yoga', 'classes', ',', 'salad', 'and', 'healthy',
, 'including', 'dental', 'insurance', ',', 'vision', 'benefits', ',', 'life', 'insurance',
['experience', ':', '2-5', 'years', 'job', 'location', ':', '-', 'aurangabad/pune', 'vacanci',
['job', 'code', ':', 'id190704', 'job', 'description', ':', 'experience', ':', '1', 'to', '2',
['job', 'description', 'summary', 'highly', 'skilled', 'technical', 'role', 'ultimate', 'resp',
['job', 'summary', 'our', 'mission', 'at', 'style', 'theory', 'is', 'to', 'build', 'a', 'sus',
['country', 'place', 'mortgage', 'founded', '1995', 'subsidiary', 'cavco', 'industry', 'inc',
['position', ':', 'wordpress', 'developer', 'location', ':', 'prahladnagar', ',', 'ahmedabad',
['monetizemore', 'builds', 'industry', 'leading', 'ad', 'technology', 'that', 'is', 'seen',
, '5', 'work', 'day', '.', 'take', 'your', 'life', 'back', 'into', 'your', 'hands', 'by', 'x',
['qualification', 'essential', 'qualification', 'related', 'ba/bs', 'two', 'year', 'work', 'e',
['job', 'summary', '*', '*', 'we', 'are', 'looking', 'for', 'fresher', 'php', 'developers',
['job', 'responsibilities', ':', 'design', 'and', 'build', 'applications', 'for', 'the', 'ios',
['software', 'developer', 'who', 'we', 'are', ':', 'the', 'american', 'physical', 'society',

['full', 'job', 'description', 'coordinating', 'with', 'development', 'teams', 'to', 'determ
 ['vidooly', 'is', 'looking', 'for', 'candidates', 'with', 'analytical', 'minds', 'and', 'resu
 ['no', 'relocation', 'needed', ',', 'can', 'telework', 'from', 'home', ',', 'required', 'goo
 ['vista', 'equity', 'partners', 'is', 'the', 'world', '"s', 'leading', 'private', 'equity',
 ['job', 'description', ':', '1-4', 'years', 'of', 'experience', 'as', 'a', 'php', 'developer
 ['1', '-', '3', 'years', 'experience', 'front-end', 'skills', ':', 'html', ',', 'css', ',',
 ['staff', 'machine', 'learning', 'scientist', 'hyderabad/bangalore', '(', 'remote', ')', 'cor
 ['summary', 'of', 'the', 'position', ':', 'we', 'are', 'seeking', 'a', 'backend', 'developer
 ['mandatory', 'skills', ':', 'kubernetes', 'and', 'docker', 'experience', 'with', 'devops',
 ['be', '/', 'bca', '/', 'any', 'bachelor', 'degree', 'certification', ':', '-', 'ccnp', ',',
 ['we', 'are', 'on', 'the', 'hunt', 'for', 'a', 'wordpress', 'developer', 'with', '2-4', 'year
 ['job', 'duties', ':', 'hardware', 'and', 'software', 'installation', 'and', 'troubleshooting
 ['job', 'description', 'we', 'are', 'looking', 'for', 'a', 'php', 'developer', 'responsible
 ['nokia', 'is', 'a', 'global', 'leader', 'in', 'the', 'technologies', 'that', 'connect', 'pee
 ['if', 'you', 'are', 'a', 'current', 'motorola', 'solutions', 'employee', ',', 'please', 'cl
 ['we', 'are', 'seeking', 'talented', 'and', 'creative', 'individuals', 'for', 'a', 'java', 'c
 ['greetings', 'of', 'the', 'day', '!!', '!!', '!!', '!!', 'position-', 'wordpress', 'developer/
 ['job', 'description', '3+', 'years', ']', 'experience', 'working', 'on', 'node/express', 'an
 ['we', 'are', 'looking', 'for', 'experienced', '&', 'well', 'knowledge', 'in', 'react', 'nat
 ['company', 'description', ':', 'hirexp', 'is', 'committed', 'to', 'building', 'collaborativ
 ['requirements', 'strong', 'background', 'in', 'linux/unix', 'administration', 'and', 'linux
 ['we', 'are', 'looking', 'for', 'an', 'ios', 'developer', 'who', 'possesses', 'a', 'passion
 ['role', 'proficiency', ':', 'level', 'of', 'proficiency', ':', 'independently', 'develops',
 ', 'software', 'engineeringoutcomes:1.', 'understand', 'and', 'provide', 'input', 'to', 'the
 ', 'software', 'engineering', 'and', 'developer', '2', '-', 'software', 'engineering', 'to',
 ', 'software', 'engineering', 'on', 'technical', 'design', '.', 'review', 'activities', 'and
 ['*', 'candidates', 'having', 'project', 'closure', 'experience', 'will', 'be', 'preferred',
 ['job', 'description', ':', 'hitachi', 'vantara', 'combines', 'technology', ',', 'intellectu
 ['java', 'developer', '*', 'summary', 'description', ':', '*', '_', '*', 'temporarily', 'worl
 ['the', 'ios', 'swift', 'developer', 'will', 'be', 'an', 'individual', 'contributor', 'in',
 ['key', 'responsibilities', ':', '-', 'working', 'in', 'a', 'cross-functional', 'team', ',',
 ['hiring', 'for', 'php-', 'laravel', 'developers', 'experience', ':', '3-5', 'years', 'locat
 ['as', 'an', 'industry', 'leader', 'and', 'software-as-a-service', 'provider', 'our', 'missi
 ['please', 'write', 'back', 'with', 'your', 'cv', 'to', 'and', 'hr', '@', 'rabsnetsolutions.
 ['description', 'job', 'department', ':', 'operations', 'job', 'title', ':', 'devops', 'engi
 ['we', 'are', 'looking', 'to', 'hire', 'a', 'skilled', 'wordpress', 'developer', 'to', 'desig
 ['amazon', 'strives', 'to', 'be', 'earth', '"s', 'most', 'customer-centric', 'company', 'whe
 ['javascript', 'developers', 'as', 'a', 'javascript', 'developer', 'you', ']', 'll', 'be', 'i
 ['experience', '5-10', 'years', 'work', 'location', 'bangalore', 'base', 'location', 'bangal
 ['product', 'operations', 'specialist', 'the', 'product', 'team', 'of', 'a', 'leading', 'cise
 ['we', 'would', 'love', 'to', 'have', 'you', 'on-board', 'if', ':', 'build', 'the', 'app', 'i
 ['have', 'two', 'or', 'more', 'ios/android', 'apps', 'developed', 'with', 'flutter', '.', 'e
 ['desired', 'skills', ':', '*', 'web', 'api', 'development', 'using', 'node.js', '.', 'exper

['software', 'engineer', ':', 'javascript', 'about', 'naehas', 'join', 'a', 'pre-ipo', ',',
 putting', 'the', 'customer', 'first', 'every', 'day', '-and', 'meaning', 'it', '.', 'being',
 ['development', 'of', 'high', 'quality', 'web', 'applications', 'like', 'online', 'communiti
 ['overview', 'cure', 'start', 'fred', 'hutchinson', 'cancer', 'research', 'center', 'home',
 ['must', 'have', 'worked', 'with', 'flutter', '/', 'android', 'application', 'development',
 ['selected', 'intern', '"s", 'day-to-day', 'responsibilities', 'include', ':', '•', 'work',
 ['overview', 'associate', 'front', 'end/full', 'stack', 'developer', 'prolifics', 'seeking',
 ['responsibilities', 'and', 'duties', '*', 'strong', 'experience', 'of', '1', 'years', 'in',
 ['company', 'description', 'klearnow', 'is', 'an', 'exciting', 'new', 'company', 'providing',
 ['punchh', 'is', 'the', 'leader', 'in', 'customer', 'loyalty', ',', 'offer', 'management', '
 ['come', 'join', 'our', 'army', 'of', 'exceptional', 'developers', 'responsible', 'for', 'de
 ['job', 'summary', 'published', 'on', ':', '9', 'months', 'ago', 'vacancy', ':', '4', 'emplo
 ['', 'small', 'team', 'experienced', 'developer', 'determined', 'change', 'modern', 'mobili
 ['snapstream', 'small', 'agile', 'tech', 'company', 'based', 'washington/heights', 'area', 'I
 ['dear', 'candidate', ',', '*', 'greetings', 'from', '*', 'tech', 'integrity', 'service', '!
 ['job', 'summary', 'we', 'are', 'looking', 'for', 'the', 'senior', 'core', 'php', 'developer
 ['highlight', 'full', 'stack', 'developer', 'ideal', 'candidate', 'responsible', 'developing
 ['database', 'administrator', '&', 'developer', 'database', 'administrator/developer', 'full
 ['responsibilities', 'develop', 'functional', 'and', 'design', 'specifications', 'for', 'serv
 ['we', 'use', 'machine', 'learning', 'at', 'the', 'very', 'core', 'of', 'the', 'platform', 't
 ['summary', 'directly', 'support', 'lending', 'management', 'facilitation', 'participation',
 ['job', 'title', ':', 'java', 'developer', 'experience', ':', 'senior', '(', '6+', 'years',
 ['wilshire', 'associate', 'incorporated', 'global', 'investment', 'advisory', 'firm', 'offer
 ['ability', 'to', 'stay', 'organized', 'while', 'multitasking', 'in', 'a', 'fast', 'paced',
 ['about', 'us', ':', 'careerninja', 'is', 'building', 'a', 'personalised', '&', 'experientia
 ['we', 'are', 'looking', 'for', 'an', 'analytical', ',', 'results-driven', 'back-end', 'devel
 ['location', ':', 'bangalore', 'experience', ':', '4', 'to', '5', 'year', 'educational', 'qua
 ['technology', ':', 'significant', 'experience', 'building', 'frontend', 'applications', '.'
 ['software/systems', 'operation', 'administrator', 'onboard', 'new', 'client', 'hosting', 'in
 ['software', 'developer', '-', 'authentication', 'remote', 'possible', 'affinity', '?', 'aff
 ['position', 'title', 'data', 'scientist', '01-oct-2020', 'job', 'id', '303875br', 'job', 'de
 ', 'both', 'globally', 'and', 'locally', '-', 'to', 'support', 'our', 'patients', '.', 'nbs'
 ['exposure', 'on', 'wordpress', 'cms', ',', 'javascript/', 'jquery', '.', 'good', 'experience
 ['vista', 'equity', 'partners', 'is', 'the', 'world', '"s", 'leading', 'private', 'equity',
 ['objective', ':', 'candidates', 'can', 'look', 'forward', 'to', 'opportunity', 'for', 'softw
 ['we', 'are', 'looking', 'for', 'a', 'machine', 'learning', '(', 'ml', ')', 'engineer', 'to'
 ['wordpress', 'cms', 'web', 'developer', 'experience', 'in', 'wordpress', 'core', ',', 'libra
 ['title', ':', 'node', 'js', 'developer', 'exp', ':', '3+', 'years', 'location', ':', 'hyder
 ['overview', 'worldventures', 'dynamic', 'global', 'member-based', 'travel', 'company', 'curr
 ['founded', 'in', '2006', 'as', 'a', 'workforce', 'strategy', 'consultancy', ',', 'brightfie
 ['we', 'are', 'looking', 'for', 'a', 'senior', 'software', 'engineer', 'to', 'join', 'our',
 ['position', ':', 'ios', 'developer', 'experience', ':', '2', '-', '4', 'years', 'location',
 ['teamwork', 'php', 'back', 'end', 'development', ',', 'maintenance', ',', 'and', 'constant'