# linkedin-job

November 29, 2024

[2]:
```
import pandas as pd
import numpy as np
```

**Loading the data and displaying :**

[3]:
```
df=pd.read_csv('/kaggle/input/linkedinjobdata/linkdin_Job_data.csv')
```

[4]:
```
df.head()
```

[4]:
```
       job_ID                                                job  \
0  3471657636  Data Analyst, Trilogy (Remote) - $60,000/year USD
1  3471669068  Data Analyst, Trilogy (Remote) - $60,000/year USD
2  3474349934                                  Data Analyst - WFH
3  3472816027                                        Data Analyst
4  3473311511                                        Data Analyst

                        location  company_id              company_name  \
0            Delhi, Delhi, India         NaN                 Crossover
1        New Delhi, Delhi, India         NaN                 Crossover
2         Greater Bengaluru Area         NaN                    Uplers
3       Gurugram, Haryana, India         NaN             PVAR SERVICES
4  Mohali district, Punjab, India         NaN  Timeline Freight Brokers

  work_type             full_time_remote  \
0    Remote         Full-time · Associate
1    Remote         Full-time · Associate
2    Remote  Full-time · Mid-Senior level
3   On-site                    Full-time
4   On-site                    Full-time

                                    no_of_employ no_of_application  \
0  1,001-5,000 employees · IT Services and IT Con…               200
1  1,001-5,000 employees · IT Services and IT Con…               184
2  1,001-5,000 employees · IT Services and IT Con…               200
3                                1-10 employees                200
4                                1-10 employees                  8

  posted_day_ago            alumni          Hiring_person  \
```

```
0        8 hours  12 company alumni                    NaN
1        8 hours  12 company alumni                    NaN
2        9 hours   3 company alumni          Shahid Ahmad
3        7 hours                NaN          Vartika Singh
4     26 minutes   1 company alumni  Manisha (Gisele Smith)

    linkedin_followers                                hiring_person_link  \
0  5,395,547 followers                                               NaN
1  5,395,547 followers                                               NaN
2                  NaN  https://www.linkedin.com/in/shahid-ahmad-a2613…
3      2,094 followers         https://www.linkedin.com/in/vartika-singh-
4                  NaN     https://www.linkedin.com/in/manisharathore0029

                                      job_details  Column1
0  About the job Crossover is the world's #1 sour…      NaN
1  About the job Crossover is the world's #1 sour…      NaN
2  About the job Profile: ML EngineersExperience:…      NaN
3  About the job Designation: Data AnalystLocatio…      NaN
4  About the job The ideal candidate will use the…      NaN
```

[5]:
```
df.tail()
```

[5]:
```
          job_ID                             job                    location  \
7922  3472039871             Back End Developer         Kochi, Kerala, India
7923  3473194471     Software Engineer - Senior     Gurugram, Haryana, India
7924  3461005032                         Vue JS  Hyderabad, Telangana, India
7925  3474305684                 iOS Developer  Bengaluru, Karnataka, India
7926  3459352801  Power train function developer  Bengaluru, Karnataka, India

      company_id               company_name work_type  \
7922         NaN           Orion Innovation    Hybrid
7923         NaN                     Uplers   On-site
7924         NaN  Tata Consultancy Services   On-site
7925         NaN                     Uplers    Remote
7926         NaN                    Akkodis    Hybrid

                 full_time_remote  \
7922         Full-time · Associate
7923  Full-time · Mid-Senior level
7924  Full-time · Mid-Senior level
7925  Full-time · Mid-Senior level
7926  Full-time · Mid-Senior level

                                       no_of_employ no_of_application  \
7922  5,001-10,000 employees · IT Services and IT Co…               25
7923  1,001-5,000 employees · IT Services and IT Con…               18
7924  10,001+ employees · IT Services and IT Consulting              15
```

```
7925  1,001-5,000 employees · IT Services and IT Con…                17
7926  10,001+ employees · IT Services and IT Consulting                 43

     posted_day_ago                 alumni         Hiring_person  \
7922         2 days      24 company alumni  Poornima Viswanathan
7923         2 days       3 company alumni         Tejveer Singh
7924         6 days  10,080 company alumni          ANNIE ANTONY
7925          1 day       3 company alumni           Arjun Jaggi
7926         1 week      27 company alumni                   NaN

        linkedin_followers                                hiring_person_link  \
7922                   NaN  https://www.linkedin.com/in/poornima-viswanath…
7923                   NaN  https://www.linkedin.com/in/tejveer-singh-a348…
7924  11,923,634 followers  https://www.linkedin.com/in/annie-antony-a1041…
7925                   NaN            https://www.linkedin.com/in/arjunjaggi7
7926   1,421,090 followers                                               NaN

                                        job_details  Column1
7922  About the job The ideal candidate will show in…      NaN
7923  About the job Experience: 4 - 8 yearsProfile: …      NaN
7924  About the job Role- Vue js DeveloperExperience…      NaN
7925  About the job Profile: iOS DeveloperExperience…      NaN
7926  About the job JD Powertrain Function Developer…      NaN
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7927 entries, 0 to 7926
Data columns (total 16 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   job_ID              7927 non-null   int64
 1   job                 7894 non-null   object
 2   location            7894 non-null   object
 3   company_id          0 non-null      float64
 4   company_name        7892 non-null   object
 5   work_type           7736 non-null   object
 6   full_time_remote    7848 non-null   object
 7   no_of_employ        7603 non-null   object
 8   no_of_application   7887 non-null   object
 9   posted_day_ago      7920 non-null   object
 10  alumni              4858 non-null   object
 11  Hiring_person       5720 non-null   object
 12  linkedin_followers  4814 non-null   object
 13  hiring_person_link  5720 non-null   object
 14  job_details         7881 non-null   object
 15  Column1             0 non-null      float64
```

```
dtypes: float64(2), int64(1), object(13)
memory usage: 991.0+ KB
```

[7]:
```
df.isnull().sum()
```

[7]:
```
job_ID                  0
job                    33
location               33
company_id           7927
company_name           35
work_type             191
full_time_remote       79
no_of_employ          324
no_of_application      40
posted_day_ago          7
alumni               3069
Hiring_person        2207
linkedin_followers   3113
hiring_person_link   2207
job_details            46
Column1              7927
dtype: int64
```

[8]:
```
df.
  ↪drop(columns=['company_id','alumni','hiring_person_link','Column1'],inplace=True)
```

[10]:
```
job=df.copy()
```

[11]:
```
job['work_type'] = job['work_type'].fillna(job['work_type'].mode()[0])
job['Hiring_person'] = job['Hiring_person'].fillna('NA')
job['linkedin_followers'] = job['linkedin_followers'].fillna('NA')
job = job.dropna(subset=['job', 'no_of_employ', 'location', 'company_name',␣
  ↪'full_time_remote', 'posted_day_ago', 'job_details'])

job.isnull().sum()
```

[11]:
```
job_ID               0
job                  0
location             0
company_name         0
work_type            0
full_time_remote     0
no_of_employ         0
no_of_application    0
posted_day_ago       0
Hiring_person        0
linkedin_followers   0
```

```
job_details           0
dtype: int64
```

**Exploratory Data Analysis :**

[12]:
```
import seaborn as sns
import matplotlib.pyplot as plt
```

[13]:
```
job.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 7583 entries, 0 to 7926
Data columns (total 12 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   job_ID              7583 non-null   int64
 1   job                 7583 non-null   object
 2   location            7583 non-null   object
 3   company_name        7583 non-null   object
 4   work_type           7583 non-null   object
 5   full_time_remote    7583 non-null   object
 6   no_of_employ        7583 non-null   object
 7   no_of_application   7583 non-null   object
 8   posted_day_ago      7583 non-null   object
 9   Hiring_person       7583 non-null   object
 10  linkedin_followers  7583 non-null   object
 11  job_details         7583 non-null   object
dtypes: int64(1), object(11)
memory usage: 770.1+ KB
```

[14]:
```
job.describe(include='all')
```

[14]:

| | job_ID | job | \ |
|---|---|---|---|
| count | 7.583000e+03 | 7583 | |
| unique | NaN | 2903 | |
| top | NaN | Lead Java Software Engineer | |
| freq | NaN | 170 | |
| mean | 3.467069e+09 | NaN | |
| std | 5.657636e+07 | NaN | |
| min | 1.419216e+08 | NaN | |
| 25% | 3.467370e+09 | NaN | |
| 50% | 3.471895e+09 | NaN | |
| 75% | 3.476185e+09 | NaN | |
| max | 3.477823e+09 | NaN | |

| | location | company_name | work_type | \ |
|---|---|---|---|---|
| count | 7583 | 7583 | 7583 | |

| | | | |
|---|---|---|---|
| unique | 139 | 2342 | 3 |
| top | Bengaluru, Karnataka, India | EPAM Anywhere | On-site |
| freq | 1268 | 1507 | 3238 |
| mean | NaN | NaN | NaN |
| std | NaN | NaN | NaN |
| min | NaN | NaN | NaN |
| 25% | NaN | NaN | NaN |
| 50% | NaN | NaN | NaN |
| 75% | NaN | NaN | NaN |
| max | NaN | NaN | NaN |

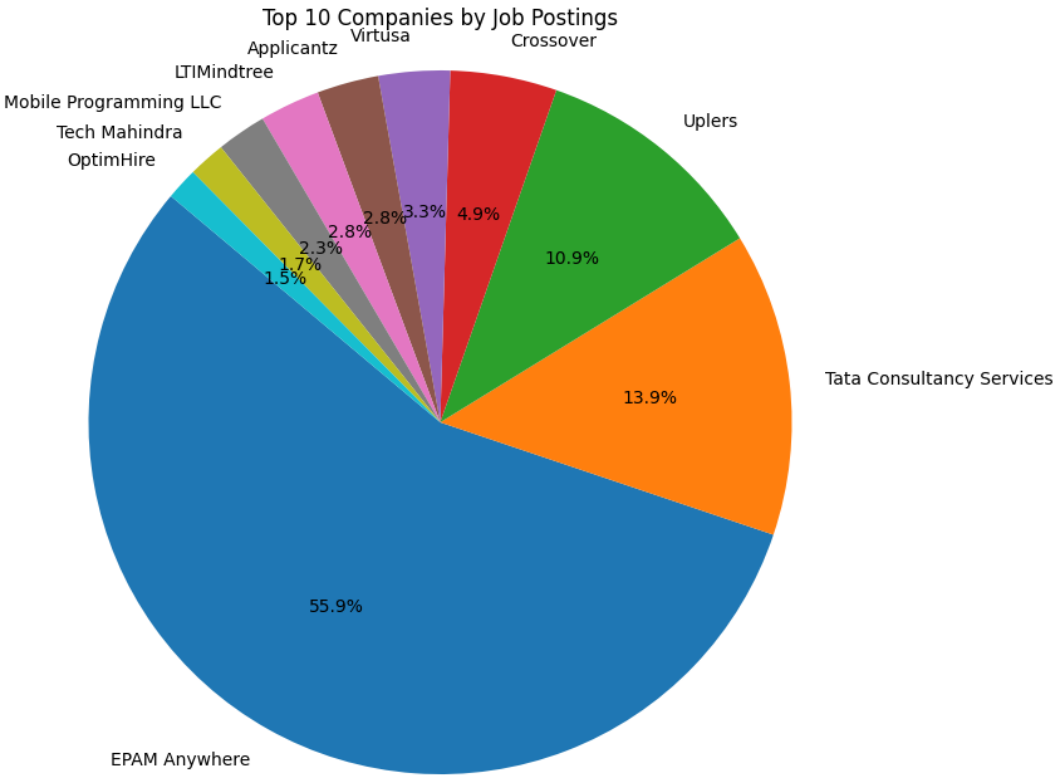
| | full_time_remote |
|---|---|
| count | 7583 |
| unique | 23 |
| top | Full-time · Mid-Senior level |
| freq | 3760 |
| mean | NaN |
| std | NaN |
| min | NaN |
| 25% | NaN |
| 50% | NaN |
| 75% | NaN |
| max | NaN |

| | no_of_employ | no_of_application |
|---|---|---|
| count | 7583 | 7583 |
| unique | 269 | 202 |
| top | 1,001-5,000 employees · IT Services and IT Con… | hours |
| freq | 2026 | 2545 |
| mean | NaN | NaN |
| std | NaN | NaN |
| min | NaN | NaN |
| 25% | NaN | NaN |
| 50% | NaN | NaN |
| 75% | NaN | NaN |
| max | NaN | NaN |

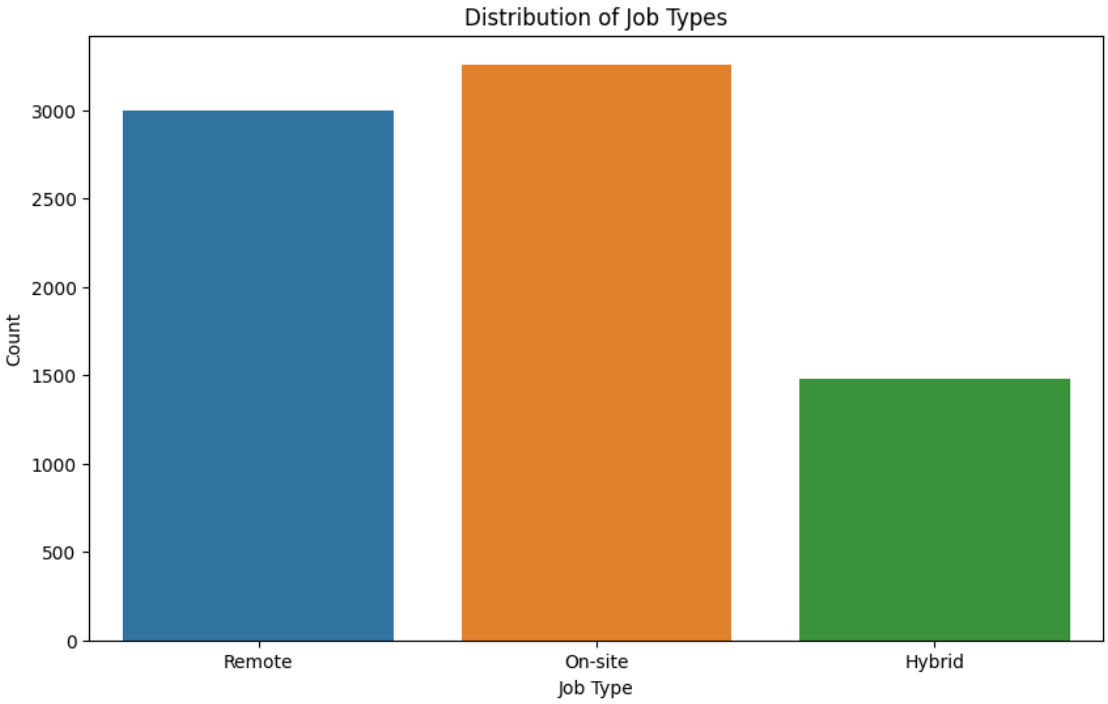
| | posted_day_ago | Hiring_person | linkedin_followers |
|---|---|---|---|
| count | 7583 | 7583 | 7583 |
| unique | 91 | 2671 | 3877 |
| top | 1 day | NA | NA |
| freq | 755 | 2123 | 2860 |
| mean | NaN | NaN | NaN |
| std | NaN | NaN | NaN |
| min | NaN | NaN | NaN |
| 25% | NaN | NaN | NaN |
| 50% | NaN | NaN | NaN |

```
75%                NaN           NaN                 NaN
max                NaN           NaN                 NaN

                                                job_details
count                                                  7583
unique                                                 4388
top      About the job Profile: Data EngineerExperience…
freq                                                     32
mean                                                    NaN
std                                                     NaN
min                                                     NaN
25%                                                     NaN
50%                                                     NaN
75%                                                     NaN
max                                                     NaN
```

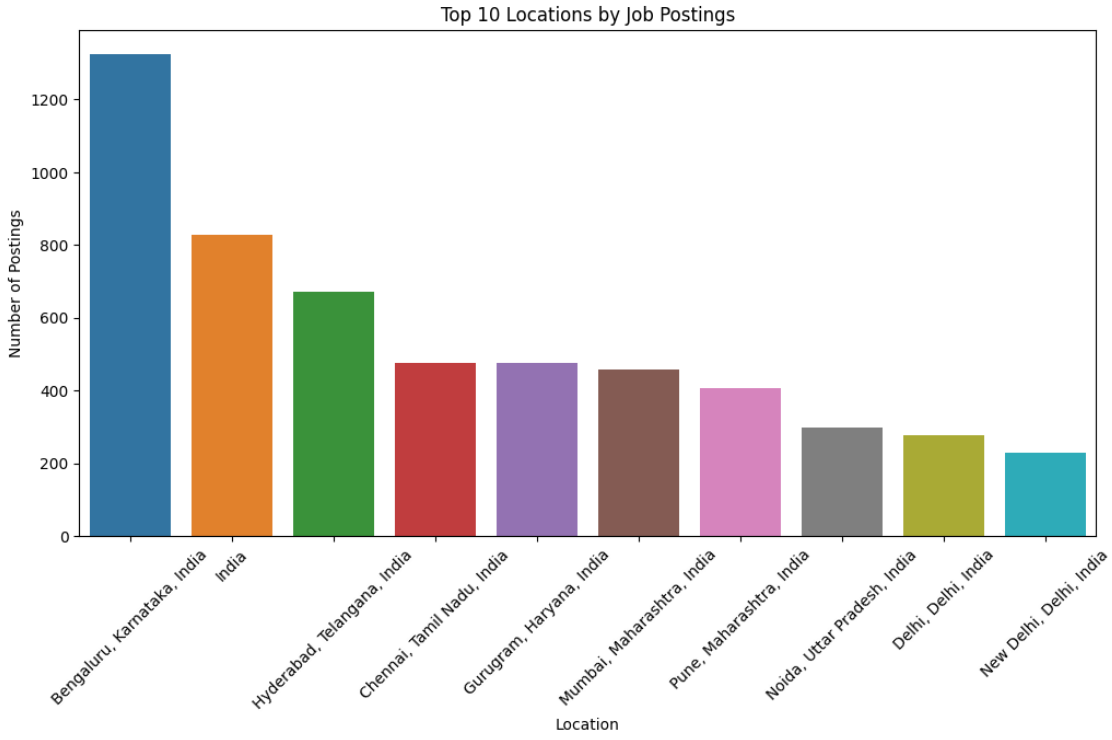
[35]:
```
company_counts = df['company_name'].value_counts()

top_10_companies = company_counts.head(10)
plt.figure(figsize=(10, 8))
plt.pie(top_10_companies, labels=top_10_companies.index, autopct='%1.1f%%',␣
 ↪startangle=140)
plt.title('Top 10 Companies by Job Postings')
plt.axis('equal')
plt.show()
```




[36]:
```
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='work_type')
plt.title('Distribution of Job Types')
plt.xlabel('Job Type')
plt.ylabel('Count')
plt.show()
```

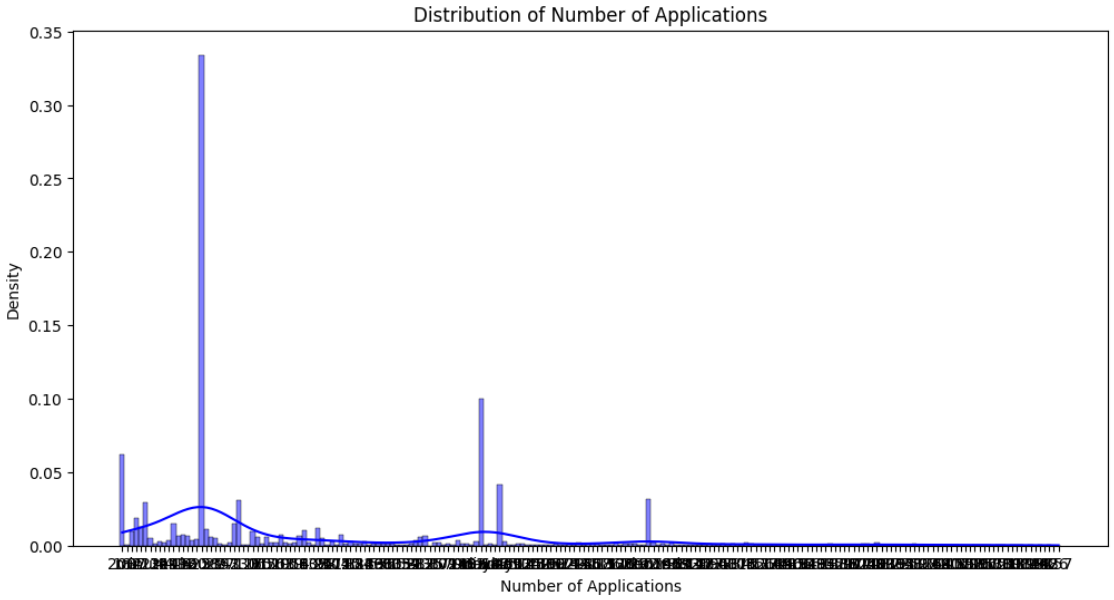



[37]:
```
top_locations = df['location'].value_counts().head(10)

plt.figure(figsize=(12, 6))
sns.barplot(x=top_locations.index, y=top_locations.values)
plt.title('Top 10 Locations by Job Postings')
plt.xlabel('Location')
plt.ylabel('Number of Postings')
plt.xticks(rotation=45)
plt.show()
```

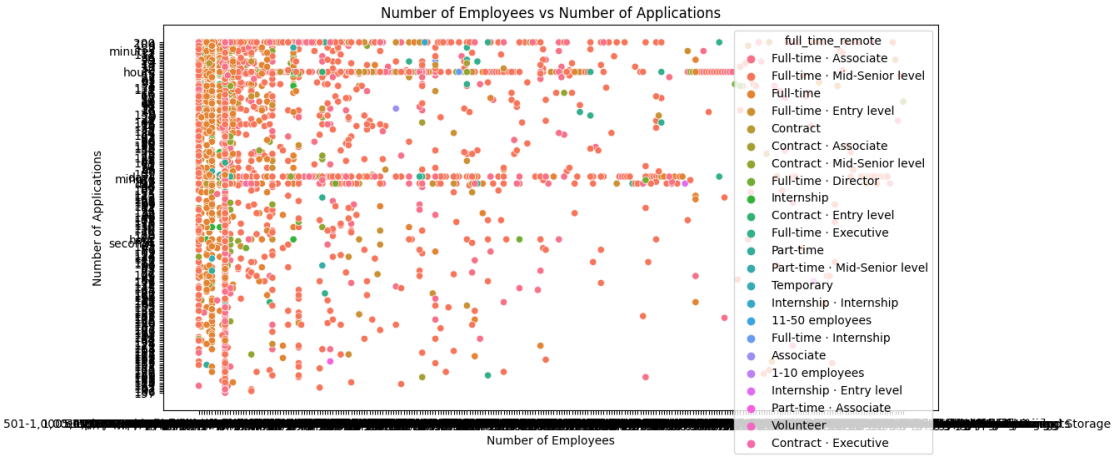



[33]:
```
plt.figure(figsize=(12,6))
sns.histplot(df['no_of_application'], bins=100, kde=True, stat="density",␣
↪color='blue', alpha=0.5)
plt.title('Distribution of Number of Applications')
plt.xlabel('Number of Applications')
plt.ylabel('Density')
plt.show()
```

```
/opt/conda/lib/python3.10/site-packages/seaborn/_oldcore.py:1119: FutureWarning:
use_inf_as_na option is deprecated and will be removed in a future version.
Convert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```




[34]:
```
plt.figure(figsize=(12, 6))
sns.scatterplot(data=df, x='no_of_employ', y='no_of_application',␣
↪hue='full_time_remote')
plt.title('Number of Employees vs Number of Applications')
plt.xlabel('Number of Employees')
plt.ylabel('Number of Applications')
plt.show()
```




[42]:
```
print(df.describe())
print(df.nunique())
```

```
              job_ID
count   7.927000e+03
mean    3.466724e+09
std     5.778011e+07
min     1.419216e+08
25%     3.467367e+09
50%     3.471882e+09
75%     3.476181e+09
max     3.477823e+09
job_ID                5843
job                   2991
location               151
company_name          2495
work_type                3
full_time_remote        23
no_of_employ           269
no_of_application      202
posted_day_ago          91
Hiring_person         2823
linkedin_followers    3935
job_details           4563
dtype: int64
```