

```
In [107]: import pandas as pd
import numpy as np
```

```
In [108]: df = pd.read_csv("C:/Users/hp/OneDrive/Desktop/survey_results_public.csv")
df.head()
```

Out[108]:

| | Responded | MainBranch | Employment | RemoteWork | CodingActivities | EdLevel | LearnCode | LearnCodeOnline | LearnCod |
|---|-----------|---------------------------------------------------|---------------------|--------------------------------------|------------------------------------------|-------------------------------------------------|---------------------------------------------------|-----------------------------------------|-----------|
| 0 | 1 | None of these | NaN | NaN | NaN | NaN | NaN | NaN | |
| 1 | 2 | I am a developer by profession | Employed, full-time | Fully remote | Hobby;Contribute to open-source projects | NaN | NaN | NaN | |
| 2 | 3 | I am not primarily a developer, but I write co... | Employed, full-time | Hybrid (some remote, some in-person) | Hobby | Master's degree (M.A., M.S., M.Eng., MBA, etc.) | Books / Physical media;Friend or family member... | documentation;Blogs;Programming Game... | Technical |
| 3 | 4 | I am a developer by profession | Employed, full-time | Fully remote | I don't code outside of work | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;School (i.e., Universit... | | NaN |
| 4 | 5 | I am a developer by profession | Employed, full-time | Hybrid (some remote, some in-person) | Hobby | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Other online resources (e.g., videos, blogs, f... | documentation;Blogs;Stack Overflow;O... | Technical |

5 rows × 79 columns

```
In [109]: df.tail()
```

Out[109]:

| | Responded | MainBranch | Employment | RemoteWork | CodingActivities | EdLevel | LearnCode | LearnCodeOnline | |
|-------|-----------|---------------------------------------------------|---------------------------------------------------|--------------------------------------|---------------------------------------------------|-------------------------------------------------|---------------------------------------------------|-----------------------------------------|-----------|
| 73263 | 73264 | I am a developer by profession | Employed, full-time | Fully remote | Freelance/contract work | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;Other online resources ... | documentation;Blogs;Written Tutorial... | Technical |
| 73264 | 73265 | I am a developer by profession | Employed, full-time | Full in-person | Hobby | Master's degree (M.A., M.S., M.Eng., MBA, etc.) | Other online resources (e.g., videos, blogs, f... | documentation;Blogs;Written Tutorial... | Technical |
| 73265 | 73266 | I am not primarily a developer, but I write co... | Employed, full-time | Hybrid (some remote, some in-person) | Hobby;School or academic work | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;Other online resources ... | documentation;Programming Games;Stac... | Uden |
| 73266 | 73267 | I am a developer by profession | Employed, full-time | Hybrid (some remote, some in-person) | Hobby | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;On the job training | | NaN |
| 73267 | 73268 | I used to be a developer by profession, but no... | Independent contractor, freelancer, or self-em... | Fully remote | Hobby;Contribute to open-source projects;Boots... | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;Friend or family member... | documentation;Blogs;Programming Game... | Technical |

5 rows × 79 columns

```
In [110]: df.describe()
```

Out[110]:

| | Responseld | CompTotal | VCHostingPersonal use | VCHostingProfessional use | WorkExp | ConvertedCompYearly |
|-------|--------------|--------------|-----------------------|---------------------------|--------------|---------------------|
| count | 73268.000000 | 3.842200e+04 | 0.0 | 0.0 | 36769.000000 | 3.807100e+04 |
| mean | 36634.500000 | 2.342434e+52 | NaN | NaN | 10.242378 | 1.707613e+05 |
| std | 21150.794099 | 4.591478e+54 | NaN | NaN | 8.706850 | 7.814132e+05 |
| min | 1.000000 | 0.000000e+00 | NaN | NaN | 0.000000 | 1.000000e+00 |
| 25% | 18317.750000 | 3.000000e+04 | NaN | NaN | 4.000000 | 3.583200e+04 |
| 50% | 36634.500000 | 7.750000e+04 | NaN | NaN | 8.000000 | 6.784500e+04 |
| 75% | 54951.250000 | 1.540000e+05 | NaN | NaN | 15.000000 | 1.200000e+05 |
| max | 73268.000000 | 9.000000e+56 | NaN | NaN | 50.000000 | 5.000000e+07 |

In [111...

check for data summary#
df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 73268 entries, 0 to 73267
Data columns (total 79 columns):
```

| # | Column | Non-Null Count | Dtype |
|----|--------------------------------|----------------|---------|
| 0 | ResponseId | 73268 non-null | int64 |
| 1 | MainBranch | 73268 non-null | object |
| 2 | Employment | 71709 non-null | object |
| 3 | RemoteWork | 58958 non-null | object |
| 4 | CodingActivities | 58899 non-null | object |
| 5 | EdLevel | 71571 non-null | object |
| 6 | LearnCode | 71580 non-null | object |
| 7 | LearnCodeOnline | 50685 non-null | object |
| 8 | LearnCodeCoursesCert | 29389 non-null | object |
| 9 | YearsCode | 71331 non-null | object |
| 10 | YearsCodePro | 51833 non-null | object |
| 11 | DevType | 61302 non-null | object |
| 12 | OrgSize | 51039 non-null | object |
| 13 | PurchaseInfluence | 50969 non-null | object |
| 14 | BuyNewTool | 67963 non-null | object |
| 15 | Country | 71771 non-null | object |
| 16 | Currency | 51264 non-null | object |
| 17 | CompTotal | 38422 non-null | float64 |
| 18 | CompFreq | 44425 non-null | object |
| 19 | LanguageHaveWorkedWith | 70975 non-null | object |
| 20 | LanguageWantToWorkWith | 67027 non-null | object |
| 21 | DatabaseHaveWorkedWith | 60121 non-null | object |
| 22 | DatabaseWantToWorkWith | 51014 non-null | object |
| 23 | PlatformHaveWorkedWith | 49924 non-null | object |
| 24 | PlatformWantToWorkWith | 40415 non-null | object |
| 25 | WebframeHaveWorkedWith | 53544 non-null | object |
| 26 | WebframeWantToWorkWith | 46122 non-null | object |
| 27 | MiscTechHaveWorkedWith | 44992 non-null | object |
| 28 | MiscTechWantToWorkWith | 36810 non-null | object |
| 29 | ToolsTechHaveWorkedWith | 54171 non-null | object |
| 30 | ToolsTechWantToWorkWith | 46566 non-null | object |
| 31 | NEWCollabToolsHaveWorkedWith | 70347 non-null | object |
| 32 | NEWCollabToolsWantToWorkWith | 64108 non-null | object |
| 33 | OpSysProfessional use | 65503 non-null | object |
| 34 | OpSysPersonal use | 70963 non-null | object |
| 35 | VersionControlSystem | 71379 non-null | object |
| 36 | VCInteraction | 68156 non-null | object |
| 37 | VCHostingPersonal use | 0 non-null | float64 |
| 38 | VCHostingProfessional use | 0 non-null | float64 |
| 39 | OfficeStackAsyncHaveWorkedWith | 46223 non-null | object |
| 40 | OfficeStackAsyncWantToWorkWith | 32072 non-null | object |
| 41 | OfficeStackSyncHaveWorkedWith | 62128 non-null | object |
| 42 | OfficeStackSyncWantToWorkWith | 47688 non-null | object |
| 43 | Blockchain | 71071 non-null | object |
| 44 | NEWSOSites | 71365 non-null | object |
| 45 | SOVisitFreq | 70961 non-null | object |
| 46 | SOAccount | 71572 non-null | object |
| 47 | SOPartFreq | 58229 non-null | object |
| 48 | SOComm | 71408 non-null | object |
| 49 | Age | 70946 non-null | object |
| 50 | Gender | 70853 non-null | object |
| 51 | Trans | 70315 non-null | object |
| 52 | Sexuality | 66565 non-null | object |
| 53 | Ethnicity | 69474 non-null | object |
| 54 | Accessibility | 67244 non-null | object |
| 55 | MentalHealth | 66447 non-null | object |
| 56 | TBranch | 52670 non-null | object |
| 57 | ICorPM | 36283 non-null | object |
| 58 | WorkExp | 36769 non-null | float64 |
| 59 | Knowledge_1 | 35804 non-null | object |
| 60 | Knowledge_2 | 34973 non-null | object |
| 61 | Knowledge_3 | 35133 non-null | object |
| 62 | Knowledge_4 | 35097 non-null | object |
| 63 | Knowledge_5 | 35014 non-null | object |
| 64 | Knowledge_6 | 34991 non-null | object |
| 65 | Knowledge_7 | 34977 non-null | object |
| 66 | Frequency_1 | 35371 non-null | object |
| 67 | Frequency_2 | 35344 non-null | object |
| 68 | Frequency_3 | 34515 non-null | object |
| 69 | TimeSearching | 36198 non-null | object |
| 70 | TimeAnswering | 36022 non-null | object |
| 71 | Onboarding | 35679 non-null | object |
| 72 | ProfessionalTech | 34906 non-null | object |
| 73 | TrueFalse_1 | 35819 non-null | object |
| 74 | TrueFalse_2 | 35715 non-null | object |
| 75 | TrueFalse_3 | 35749 non-null | object |
| 76 | SurveyLength | 70444 non-null | object |
| 77 | SurveyEase | 70508 non-null | object |
| 78 | ConvertedCompYearly | 38071 non-null | float64 |

```
dtypes: float64(5), int64(1), object(73)
```

```
memory usage: 44.2+ MB
```

```
In [112]: df.isna().sum()
```

Out[112]:

| | |
|---------------------|-------|
| ResponseId | 0 |
| MainBranch | 0 |
| Employment | 1559 |
| RemoteWork | 14310 |
| CodingActivities | 14369 |
| ... | |
| TrueFalse_2 | 37553 |
| TrueFalse_3 | 37519 |
| SurveyLength | 2824 |
| SurveyEase | 2760 |
| ConvertedCompYearly | 35197 |

Length: 79, dtype: int64

```
In [113]: # Drop columns with a high number of missing values
threshold = 0.7
df = df.dropna(thresh=threshold * len(df), axis=1)
df.head()
```

Out[113]:

| | ResponseId | MainBranch | Employment | RemoteWork | CodingActivities | EdLevel | LearnCode | YearsCode | YearsCodePro | DevType |
|---|------------|---------------------------------------------------|---------------------|--------------------------------------|------------------------------------------|-------------------------------------------------|---------------------------------------------------|-----------|--------------|---------------------------------------------------|
| 0 | 1 | None of these | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 1 | 2 | I am a developer by profession | Employed, full-time | Fully remote | Hobby;Contribute to open-source projects | NaN | NaN | NaN | NaN | NaN |
| 2 | 3 | I am not primarily a developer, but I write co... | Employed, full-time | Hybrid (some remote, some in-person) | Hobby | Master's degree (M.A., M.S., M.Eng., MBA, etc.) | Books / Physical media;Friend or family member... | 14 | 5 | Data scientist or machine learning specialist;... |
| 3 | 4 | I am a developer by profession | Employed, full-time | Fully remote | I don't code outside of work | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;School (i.e., Universit... | 20 | 17 | Developer, full-stack |
| 4 | 5 | I am a developer by profession | Employed, full-time | Hybrid (some remote, some in-person) | Hobby | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Other online resources (e.g., videos, blogs, f... | 8 | 3 | Developer, front-end;Developer, full-stack;Dev... |

5 rows × 40 columns

```
In [114]: # Drop rows with missing values in specific columns
cols_to_drop_na_rows = ['EdLevel', 'YearsCode']
df = df.dropna(subset=cols_to_drop_na_rows)
df.head()
```

| Out[114]: | ResponseId | MainBranch | Employment | RemoteWork | CodingActivities | EdLevel | LearnCode | YearsCode | YearsCodePro | DevType |
|-----------|------------|---------------------------------------------------|---------------------|--------------------------------------|------------------------------|---------------------------------------------------|---------------------------------------------------|-----------|--------------|---------------------------------------------------|
| 2 | 3 | I am not primarily a developer, but I write co... | Employed, full-time | Hybrid (some remote, some in-person) | Hobby | Master's degree (M.A., M.S., M.Eng., MBA, etc.) | Books / Physical media;Friend or family member... | 14 | 5 | Data scientist or machine learning specialist;... |
| 3 | 4 | I am a developer by profession | Employed, full-time | Fully remote | I don't code outside of work | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;School (i.e., Universit... | 20 | 17 | Developer, full-stack |
| 4 | 5 | I am a developer by profession | Employed, full-time | Hybrid (some remote, some in-person) | Hobby | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Other online resources (e.g., videos, blogs, f... | 8 | 3 | Developer, front-end;Developer, full-stack;Dev... |
| 5 | 6 | I am not primarily a developer, but I write co... | Student, full-time | NaN | NaN | Master's degree (M.A., M.S., M.Eng., MBA, etc.) | Books / Physical media;School (i.e., Universit... | 15 | NaN | NaN |
| 6 | 7 | I code primarily as a hobby | Student, part-time | NaN | NaN | Secondary school (e.g. American high school, G... | Other online resources (e.g., videos, blogs, f... | 3 | NaN | NaN |

5 rows × 40 columns

```
In [115]: # Convert categorical columns to appropriate data types
categorical_cols = ['MainBranch', 'Employment', 'RemoteWork', 'CodingActivities', 'EdLevel']
df[categorical_cols] = df[categorical_cols].astype('category')

df.head()
```

| Out[115]: | ResponseId | MainBranch | Employment | RemoteWork | CodingActivities | EdLevel | LearnCode | YearsCode | YearsCodePro | DevType |
|-----------|------------|---------------------------------------------------|---------------------|--------------------------------------|------------------------------|---------------------------------------------------|---------------------------------------------------|-----------|--------------|---------------------------------------------------|
| 2 | 3 | I am not primarily a developer, but I write co... | Employed, full-time | Hybrid (some remote, some in-person) | Hobby | Master's degree (M.A., M.S., M.Eng., MBA, etc.) | Books / Physical media;Friend or family member... | 14 | 5 | Data scientist or machine learning specialist;... |
| 3 | 4 | I am a developer by profession | Employed, full-time | Fully remote | I don't code outside of work | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;School (i.e., Universit... | 20 | 17 | Developer, full-stack |
| 4 | 5 | I am a developer by profession | Employed, full-time | Hybrid (some remote, some in-person) | Hobby | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Other online resources (e.g., videos, blogs, f... | 8 | 3 | Developer, front-end;Developer, full-stack;Dev... |
| 5 | 6 | I am not primarily a developer, but I write co... | Student, full-time | NaN | NaN | Master's degree (M.A., M.S., M.Eng., MBA, etc.) | Books / Physical media;School (i.e., Universit... | 15 | NaN | NaN |
| 6 | 7 | I code primarily as a hobby | Student, part-time | NaN | NaN | Secondary school (e.g. American high school, G... | Other online resources (e.g., videos, blogs, f... | 3 | NaN | NaN |

5 rows × 40 columns

```
In [116]: # Handle missing values in numeric columns
numeric_cols = ['YearsCode']
df[numeric_cols] = df[numeric_cols].apply(pd.to_numeric, errors='coerce')

df.head()
```

| Out[116]: | Respondel | MainBranch | Employment | RemoteWork | CodingActivities | EdLevel | LearnCode | YearsCode | YearsCodePro | DevType |
|-----------|-----------|---------------------------------------------------|---------------------|--------------------------------------|------------------------------|---------------------------------------------------|---------------------------------------------------|-----------|--------------|---------------------------------------------------|
| 2 | 3 | I am not primarily a developer, but I write co... | Employed, full-time | Hybrid (some remote, some in-person) | Hobby | Master's degree (M.A., M.S., M.Eng., MBA, etc.) | Books / Physical media;Friend or family member... | 14.0 | 5 | Data scientist or machine learning specialist;... |
| 3 | 4 | I am a developer by profession | Employed, full-time | Fully remote | I don't code outside of work | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;School (i.e., Universit... | 20.0 | 17 | Developer, full-stack |
| 4 | 5 | I am a developer by profession | Employed, full-time | Hybrid (some remote, some in-person) | Hobby | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Other online resources (e.g., videos, blogs, f... | 8.0 | 3 | Developer, front-end;Developer, full-stack;Dev... |
| 5 | 6 | I am not primarily a developer, but I write co... | Student, full-time | NaN | NaN | Master's degree (M.A., M.S., M.Eng., MBA, etc.) | Books / Physical media;School (i.e., Universit... | 15.0 | NaN | NaN |
| 6 | 7 | I code primarily as a hobby | Student, part-time | NaN | NaN | Secondary school (e.g. American high school, G... | Other online resources (e.g., videos, blogs, f... | 3.0 | NaN | NaN |

5 rows × 40 columns

| | | | | | | | | | | |
|---|--|--|--|--|--|--|--|--|--|--|
| 4 | | | | | | | | | | |
|---|--|--|--|--|--|--|--|--|--|--|

In [117...

df.tail()

| Out[117]: | Respondel | MainBranch | Employment | RemoteWork | CodingActivities | EdLevel | LearnCode | YearsCode | YearsCodePro | Dev |
|-----------|-----------|---------------------------------------------------|---------------------------------------------------|--------------------------------------|---------------------------------------------------|-------------------------------------------------|---------------------------------------------------|-----------|--------------|--------------------------------|
| 73263 | 73264 | I am a developer by profession | Employed, full-time | Fully remote | Freelance/contract work | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;Other online resources ... | 8.0 | 5 | Deve bac |
| 73264 | 73265 | I am a developer by profession | Employed, full-time | Full in-person | Hobby | Master's degree (M.A., M.S., M.Eng., MBA, etc.) | Other online resources (e.g., videos, blogs, f... | 6.0 | 5 | Data scier m le spe |
| 73265 | 73266 | I am not primarily a developer, but I write co... | Employed, full-time | Hybrid (some remote, some in-person) | Hobby;School or academic work | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;Other online resources ... | 42.0 | 33 | Develope stack;Deve desktop c |
| 73266 | 73267 | I am a developer by profession | Employed, full-time | Hybrid (some remote, some in-person) | Hobby | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;On the job training | 50.0 | 31 | Developer, end;Deve desktop oi |
| 73267 | 73268 | I used to be a developer by profession, but no... | Independent contractor, freelancer, or self-em... | Fully remote | Hobby;Contribute to open-source projects;Boots... | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;Friend or family member... | 16.0 | 5 | Developer, end;Eng data;Eng |

5 rows × 40 columns

| | | | | | | | | | | |
|---|--|--|--|--|--|--|--|--|--|--|
| 4 | | | | | | | | | | |
|---|--|--|--|--|--|--|--|--|--|--|

In [118...

df = df.fillna(method='bfill')

In [119...

df

Out[119]:

| | ResponseId | MainBranch | Employment | RemoteWork | CodingActivities | EdLevel | LearnCode | YearsCode | YearsCodePro | De |
|-------|------------|---------------------------------------------------|---------------------------------------------------|--------------------------------------|---------------------------------------------------|---------------------------------------------------|---------------------------------------------------|-----------|--------------|------------------------------------|
| 2 | 3 | I am not primarily a developer, but I write co... | Employed, full-time | Hybrid (some remote, some in-person) | Hobby | Master's degree (M.A., M.S., M.Eng., MBA, etc.) | Books / Physical media;Friend or family member... | 14.0 | 5 | Data scie m le spec |
| 3 | 4 | I am a developer by profession | Employed, full-time | Fully remote | I don't code outside of work | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;School (i.e., Universit... | 20.0 | 17 | Develop |
| 4 | 5 | I am a developer by profession | Employed, full-time | Hybrid (some remote, some in-person) | Hobby | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Other online resources (e.g., videos, blogs, f... | 8.0 | 3 | Developer end;Dev full-stack |
| 5 | 6 | I am not primarily a developer, but I write co... | Student, full-time | Hybrid (some remote, some in-person) | I don't code outside of work | Master's degree (M.A., M.S., M.Eng., MBA, etc.) | Books / Physical media;School (i.e., Universit... | 15.0 | 6 | Develop stack;S |
| 6 | 7 | I code primarily as a hobby | Student, part-time | Hybrid (some remote, some in-person) | I don't code outside of work | Secondary school (e.g. American high school, G... | Other online resources (e.g., videos, blogs, f... | 3.0 | 6 | Develop stack;S |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 73263 | 73264 | I am a developer by profession | Employed, full-time | Fully remote | Freelance/contract work | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;Other online resources ... | 8.0 | 5 | Dev ba |
| 73264 | 73265 | I am a developer by profession | Employed, full-time | Full in-person | Hobby | Master's degree (M.A., M.S., M.Eng., MBA, etc.) | Other online resources (e.g., videos, blogs, f... | 6.0 | 5 | Data scie m le sp |
| 73265 | 73266 | I am not primarily a developer, but I write co... | Employed, full-time | Hybrid (some remote, some in-person) | Hobby;School or academic work | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;Other online resources ... | 42.0 | 33 | Develop stack;Dev desktop |
| 73266 | 73267 | I am a developer by profession | Employed, full-time | Hybrid (some remote, some in-person) | Hobby | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;On the job training | 50.0 | 31 | Developer end;Dev desktop c |
| 73267 | 73268 | I used to be a developer by profession, but no... | Independent contractor, freelancer, or self-em... | Fully remote | Hobby;Contribute to open-source projects;Boots... | Bachelor's degree (B.A., B.S., B.Eng., etc.) | Books / Physical media;Friend or family member... | 16.0 | 5 | Developer end;En data;En |

71204 rows × 40 columns

In [120]: df.columns

```
Out[120]: Index(['ResponseId', 'MainBranch', 'Employment', 'RemoteWork',
      'CodingActivities', 'EdLevel', 'LearnCode', 'YearsCode', 'YearsCodePro',
      'DevType', 'BuyNewTool', 'Country', 'LanguageHaveWorkedWith',
      'LanguageWantToWorkWith', 'DatabaseHaveWorkedWith',
      'WebframeHaveWorkedWith', 'ToolsTechHaveWorkedWith',
      'NEWCollabToolsHaveWorkedWith', 'NEWCollabToolsWantToWorkWith',
      'OpSysProfessional use', 'OpSysPersonal use', 'VersionControlSystem',
      'VCInteraction', 'OfficeStackSyncHaveWorkedWith', 'Blockchain',
      'NEWSOSites', 'SOVisitFreq', 'SOAccount', 'SOPartFreq', 'SOComm', 'Age',
      'Gender', 'Trans', 'Sexuality', 'Ethnicity', 'Accessibility',
      'MentalHealth', 'TBranch', 'SurveyLength', 'SurveyEase'],
      dtype='object')
```

```
In [121]: data = {
      'YearsCodePro': [14.0, 20.0, 8.0, 15.0, 3.0],
      'DevType': [
        'Data scientist or machine learning specialist;Other (please specify):',
        'Developer, full-stack;Other (please specify):',
```

```

        'Developer, front-end;Developer, full-stack;Developer, desktop or enterprise applications',
        'Developer, full-stack;Student;Other (please specify):',
        'Developer, full-stack;Student',
    ],
    'BuyNewTool': [
        'Multiple times per day',
        'A few times per week',
        'Daily or almost daily',
        'Multiple times per day',
        'Daily or almost daily',
    ],
    # ... (Continue the same pattern for other columns)
}

df = pd.DataFrame(data)

# Convert 'YearsCodePro' to numeric
df['YearsCodePro'] = pd.to_numeric(df['YearsCodePro'], errors='coerce')

# Remove leading and trailing whitespaces from string columns
df = df.apply(lambda x: x.str.strip() if x.dtype == "object" else x)

# Handle 'DevType' column, split the values into separate columns
dev_type_columns = df['DevType'].str.get_dummies(sep=';')
df = pd.concat([df, dev_type_columns], axis=1)

# Drop the original 'DevType' column
df.drop(columns=['DevType'], inplace=True)

# Print the cleaned DataFrame
print(df)

```

```

YearsCodePro      BuyNewTool \
0          14.0  Multiple times per day
1          20.0    A few times per week
2           8.0   Daily or almost daily
3          15.0  Multiple times per day
4           3.0   Daily or almost daily

Data scientist or machine learning specialist \
0                      1
1                      0
2                      0
3                      0
4                      0

Developer, desktop or enterprise applications  Developer, front-end \
0                      0                      0
1                      0                      0
2                      1                      1
3                      0                      0
4                      0                      0

Developer, full-stack  Other (please specify):  Student
0                      0                      1      0
1                      1                      1      0
2                      1                      0      0
3                      1                      1      1
4                      1                      0      1

```

In [126]: df.head()

```

Out[126]:
   YearsCodePro  BuyNewTool  Data scientist or machine learning specialist  Developer, desktop or enterprise applications  Developer, front-end  Developer, full-stack  Other (please specify):  Student
0          14.0  Multiple times per day                                1                                0                0                0                1            0
1          20.0    A few times per week                                0                                0                0                1                1            0
2           8.0   Daily or almost daily                                0                                1                1                0                0
3          15.0  Multiple times per day                                0                                0                0                1                1            1
4           3.0   Daily or almost daily                                0                                0                0                1                0            1

```

```

#Encoding
data = {
    'YearsCodePro': [14.0, 20.0, 8.0, 15.0, 3.0],
    'DevType': [
        'Data scientist or machine learning specialist;Other (please specify):',
        'Developer, full-stack;Other (please specify):',
        'Developer, front-end;Developer, full-stack;Developer, desktop or enterprise applications',
        'Developer, full-stack;Student;Other (please specify):',
        'Developer, full-stack;Student',
    ],
}

```



```

    'BuyNewTool': [
        'Multiple times per day',
        'A few times per week',
        'Daily or almost daily',
        'Multiple times per day',
        'Daily or almost daily',
    ],
    # ... (Continue the same pattern for other columns)
}

df = pd.DataFrame(data)

# Convert 'YearsCodePro' to numeric
df['YearsCodePro'] = pd.to_numeric(df['YearsCodePro'], errors='coerce')

# Remove leading and trailing whitespaces from string columns
df = df.apply(lambda x: x.str.strip() if x.dtype == "object" else x)

# Handle 'DevType' column, split the values into separate columns
dev_type_columns = df['DevType'].str.get_dummies(sep=';')
df = pd.concat([df, dev_type_columns], axis=1)

# Drop the original 'DevType' column
df.drop(columns=['DevType'], inplace=True)

# Print the cleaned DataFrame
print(df)

```

```

YearsCodePro      BuyNewTool \
0      14.0  Multiple times per day
1      20.0    A few times per week
2       8.0   Daily or almost daily
3      15.0  Multiple times per day
4       3.0   Daily or almost daily

Data scientist or machine learning specialist \
0      1
1      0
2      0
3      0
4      0

Developer, desktop or enterprise applications  Developer, front-end \
0      0      0
1      0      0
2      1      1
3      0      0
4      0      0

Developer, full-stack  Other (please specify):  Student
0      0      1      0
1      1      1      0
2      1      0      0
3      1      1      1
4      1      0      1

```

In [135]: df.head()

```

Out[135]:
   YearsCodePro  BuyNewTool  Data scientist or machine learning specialist  Developer, desktop or enterprise applications  Developer, front-end  Developer, full-stack  Other (please specify):  Student
0      14.0  Multiple times per day      1      0      0      0      1      0
1      20.0    A few times per week      0      0      0      1      1      0
2       8.0   Daily or almost daily      0      1      1      1      0      0
3      15.0  Multiple times per day      0      0      0      1      1      1
4       3.0   Daily or almost daily      0      0      0      1      0      1

```

```

# Existing data
existing_data = {
    'YearsCodePro': [14.0, 20.0, 8.0, 15.0, 3.0],
    'BuyNewTool': ['Multiple times per day', 'A few times per week', 'Daily or almost daily', 'Multiple times p
    'SOPartFreq': ['Neutral', 'Yes, definitely', 'Yes, definitely', 'Yes, definitely', 'Yes, definitely'],
    'SOComm': ['25-34 years old', '35-44 years old', '25-34 years old', '25-34 years old', 'Under 18 years old']
    'Age': [25, 35, 25, 25, 17],
    'Gender': ['Man', 'Man', 'Or, in your own words:', 'Or, in your own words:', 'Man'],
    'Trans': ['No', 'No', 'No', 'No', 'No'],
    'Sexuality': ['Bisexual', 'Straight / Heterosexual', 'Prefer to self-describe:', 'Prefer to self-describe:'
    'Ethnicity': ['White', 'White', 'Not specified', 'Not specified', 'Indian'],
    'Accessibility': ['None of the above', 'None of the above', 'None of the above', 'None of the above', 'None
    'MentalHealth': ['I have a mood or emotional disorder (e.g., depression, anxiety, bipolar disorder)', 'No',
    'Data scientist or machine learning specialist': [1, 0, 0, 0, 1],

```

```

    'Developer, desktop or enterprise applications': [0, 0, 0, 1, 0],
    'Developer, front-end': [0, 0, 1, 0, 0],
    'Developer, full-stack': [0, 1, 1, 1, 1],
    'Other (please specify)': [1, 0, 0, 0, 0],
    'Student': [0, 0, 0, 1, 1],
    'TBranch': ['No', 'No', 'No', 'Yes', 'Yes'],
}

existing_df = pd.DataFrame(existing_data)

# New data
new_data = {
    'YearsCodePro': [10.0, 22.0, 5.0],
    'BuyNewTool': ['Daily or almost daily', 'Once a week', 'Multiple times per day'],
    'SOPartFreq': ['Neutral', 'Neutral', 'Yes, definitely'],
    'SOComm': ['25-34 years old', '35-44 years old', '25-34 years old'],
    'Age': [30, 40, 30],
    'Gender': ['Woman', 'Man', 'Man'],
    'Trans': ['Yes', 'No', 'No'],
    'Sexuality': ['Straight / Heterosexual', 'Bisexual', 'Prefer to self-describe:'],
    'Ethnicity': ['Asian', 'Black', 'White'],
    'Accessibility': ['None of the above', 'None of the above', 'None of the above'],
    'MentalHealth': ['No', 'Yes', 'Yes'],
    'Data scientist or machine learning specialist': [0, 1, 1],
    'Developer, desktop or enterprise applications': [1, 0, 0],
    'Developer, front-end': [0, 0, 1],
    'Developer, full-stack': [1, 1, 1],
    'Other (please specify)': [0, 0, 1],
    'Student': [1, 0, 0],
    'TBranch': ['Yes', 'No', 'No'],
}

new_df = pd.DataFrame(new_data)

# Concatenate existing and new data
df = pd.concat([existing_df, new_df], ignore_index=True)

# Reorder columns with 'TBranch' at the end
column_order = [col for col in df.columns if col != 'TBranch'] + ['TBranch']
df = df[column_order]

# Print the combined DataFrame
print(df)

```

| | YearsCodePro | BuyNewTool | SOPartFreq | SOCComm | \ |
|---|--------------|------------------------|-----------------|--------------------|---|
| 0 | 14.0 | Multiple times per day | Neutral | 25-34 years old | |
| 1 | 20.0 | A few times per week | Yes, definitely | 35-44 years old | |
| 2 | 8.0 | Daily or almost daily | Yes, definitely | 25-34 years old | |
| 3 | 15.0 | Multiple times per day | Yes, definitely | 25-34 years old | |
| 4 | 3.0 | Daily or almost daily | Yes, definitely | Under 18 years old | |
| 5 | 10.0 | Daily or almost daily | Neutral | 25-34 years old | |
| 6 | 22.0 | Once a week | Neutral | 35-44 years old | |
| 7 | 5.0 | Multiple times per day | Yes, definitely | 25-34 years old | |

| | Age | Gender | Trans | Sexuality | Ethnicity | \ |
|---|-----|------------------------|-------|--------------------------|---------------|---|
| 0 | 25 | Man | No | Bisexual | White | |
| 1 | 35 | Man | No | Straight / Heterosexual | White | |
| 2 | 25 | Or, in your own words: | No | Prefer to self-describe: | Not specified | |
| 3 | 25 | Or, in your own words: | No | Prefer to self-describe: | Not specified | |
| 4 | 17 | Man | No | Straight / Heterosexual | Indian | |
| 5 | 30 | Woman | Yes | Straight / Heterosexual | Asian | |
| 6 | 40 | Man | No | Bisexual | Black | |
| 7 | 30 | Man | No | Prefer to self-describe: | White | |

| | Accessibility | MentalHealth | \ |
|---|-------------------|---------------------------------------------------|---|
| 0 | None of the above | I have a mood or emotional disorder (e.g., dep... | |
| 1 | None of the above | No | |
| 2 | None of the above | No | |
| 3 | None of the above | Yes | |
| 4 | None of the above | No | |
| 5 | None of the above | No | |
| 6 | None of the above | Yes | |
| 7 | None of the above | Yes | |

| | Data scientist or machine learning specialist | \ |
|---|-----------------------------------------------|---|
| 0 | 1 | |
| 1 | 0 | |
| 2 | 0 | |
| 3 | 0 | |
| 4 | 1 | |
| 5 | 0 | |
| 6 | 1 | |
| 7 | 1 | |

| | Developer, desktop or enterprise applications | Developer, front-end | \ |
|---|-----------------------------------------------|----------------------|---|
| 0 | 0 | 0 | |
| 1 | 0 | 0 | |
| 2 | 0 | 1 | |
| 3 | 1 | 0 | |
| 4 | 0 | 0 | |
| 5 | 1 | 0 | |
| 6 | 0 | 0 | |
| 7 | 0 | 1 | |

| | Developer, full-stack | Other (please specify): | Student | TBranch | |
|---|-----------------------|-------------------------|---------|---------|--|
| 0 | 0 | 1 | 0 | No | |
| 1 | 1 | 0 | 0 | No | |
| 2 | 1 | 0 | 0 | No | |
| 3 | 1 | 0 | 1 | Yes | |
| 4 | 1 | 0 | 1 | Yes | |
| 5 | 1 | 0 | 1 | Yes | |
| 6 | 1 | 0 | 0 | No | |
| 7 | 1 | 1 | 0 | No | |

In [137]: df.head()

Out[137]:

| | YearsCodePro | BuyNewTool | SOPartFreq | SOCComm | Age | Gender | Trans | Sexuality | Ethnicity | Accessibility | MentalHealth | Data scientist or machine learning specialist |
|---|--------------|------------------------|-----------------|--------------------|-----|------------------------|-------|--------------------------|---------------|-------------------|---------------------------------------------------|-----------------------------------------------|
| 0 | 14.0 | Multiple times per day | Neutral | 25-34 years old | 25 | Man | No | Bisexual | White | None of the above | I have a mood or emotional disorder (e.g., dep... | 1 |
| 1 | 20.0 | A few times per week | Yes, definitely | 35-44 years old | 35 | Man | No | Straight / Heterosexual | White | None of the above | No | 0 |
| 2 | 8.0 | Daily or almost daily | Yes, definitely | 25-34 years old | 25 | Or, in your own words: | No | Prefer to self-describe: | Not specified | None of the above | No | 0 |
| 3 | 15.0 | Multiple times per day | Yes, definitely | 25-34 years old | 25 | Or, in your own words: | No | Prefer to self-describe: | Not specified | None of the above | Yes | 0 |
| 4 | 3.0 | Daily or almost daily | Yes, definitely | Under 18 years old | 17 | Man | No | Straight / Heterosexual | Indian | None of the above | No | 1 |

```
In [139.. df.drop(columns=["SOComm"], inplace=True)
```

```
In [140.. df
```

Out[140]:

| | YearsCodePro | BuyNewTool | SOPartFreq | Age | Gender | Trans | Sexuality | Ethnicity | Accessibility | MentalHealth | Data scientist or machine learning specialist | Developer, desktop or enterprise applications |
|---|--------------|------------------------|-----------------|-----|------------------------|-------|--------------------------|---------------|-------------------|---------------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| 0 | 14.0 | Multiple times per day | Neutral | 25 | Man | No | Bisexual | White | None of the above | I have a mood or emotional disorder (e.g., dep... | 1 | 0 |
| 1 | 20.0 | A few times per week | Yes, definitely | 35 | Man | No | Straight / Heterosexual | White | None of the above | No | 0 | 0 |
| 2 | 8.0 | Daily or almost daily | Yes, definitely | 25 | Or, in your own words: | No | Prefer to self-describe: | Not specified | None of the above | No | 0 | 0 |
| 3 | 15.0 | Multiple times per day | Yes, definitely | 25 | Or, in your own words: | No | Prefer to self-describe: | Not specified | None of the above | Yes | 0 | 1 |
| 4 | 3.0 | Daily or almost daily | Yes, definitely | 17 | Man | No | Straight / Heterosexual | Indian | None of the above | No | 1 | 0 |
| 5 | 10.0 | Daily or almost daily | Neutral | 30 | Woman | Yes | Straight / Heterosexual | Asian | None of the above | No | 0 | 1 |
| 6 | 22.0 | Once a week | Neutral | 40 | Man | No | Bisexual | Black | None of the above | Yes | 1 | 0 |
| 7 | 5.0 | Multiple times per day | Yes, definitely | 30 | Man | No | Prefer to self-describe: | White | None of the above | Yes | 1 | 0 |

```
In [142.. columns_to_backfill = ['Age', 'MentalHealth']
df[columns_to_backfill] = df[columns_to_backfill].fillna(method='bfill')
```

```
In [143.. df
```

Out[143]:

| | YearsCodePro | BuyNewTool | SOPartFreq | Age | Gender | Trans | Sexuality | Ethnicity | Accessibility | MentalHealth | Data scientist or machine learning specialist | Developer, desktop or enterprise applications |
|---|--------------|------------------------|-----------------|-----|------------------------|-------|--------------------------|---------------|-------------------|---------------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| 0 | 14.0 | Multiple times per day | Neutral | 25 | Man | No | Bisexual | White | None of the above | I have a mood or emotional disorder (e.g., dep... | 1 | 0 |
| 1 | 20.0 | A few times per week | Yes, definitely | 35 | Man | No | Straight / Heterosexual | White | None of the above | No | 0 | 0 |
| 2 | 8.0 | Daily or almost daily | Yes, definitely | 25 | Or, in your own words: | No | Prefer to self-describe: | Not specified | None of the above | No | 0 | 0 |
| 3 | 15.0 | Multiple times per day | Yes, definitely | 25 | Or, in your own words: | No | Prefer to self-describe: | Not specified | None of the above | Yes | 0 | 1 |
| 4 | 3.0 | Daily or almost daily | Yes, definitely | 17 | Man | No | Straight / Heterosexual | Indian | None of the above | No | 1 | 0 |
| 5 | 10.0 | Daily or almost daily | Neutral | 30 | Woman | Yes | Straight / Heterosexual | Asian | None of the above | No | 0 | 1 |
| 6 | 22.0 | Once a week | Neutral | 40 | Man | No | Bisexual | Black | None of the above | Yes | 1 | 0 |
| 7 | 5.0 | Multiple times per day | Yes, definitely | 30 | Man | No | Prefer to self-describe: | White | None of the above | Yes | 1 | 0 |

```
In [166.. df['Gender'].replace({'Or, in your own words:': 'Man'}, inplace=True)

df
```

Out[166]:

| | YearsCodePro | BuyNewTool | SOPartFreq | Age | Gender | Trans | Sexuality | Ethnicity | Accessibility | MentalHealth | Data scientist or machine learning specialist | Developer, desktop or enterprise applications |
|---|--------------|------------------------|-----------------|-----|--------|-------|--------------------------|---------------|-------------------|---------------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| 0 | 14.0 | Multiple times per day | Neutral | 25 | Man | No | Bisexual | White | None of the above | I have a mood or emotional disorder (e.g., dep... | 1 | 0 |
| 1 | 20.0 | A few times per week | Yes, definitely | 35 | Man | No | Straight / Heterosexual | White | None of the above | No | 0 | 0 |
| 2 | 8.0 | Daily or almost daily | Yes, definitely | 25 | Man | No | Prefer to self-describe: | Not specified | None of the above | No | 0 | 0 |
| 3 | 15.0 | Multiple times per day | Yes, definitely | 25 | Man | No | Prefer to self-describe: | Not specified | None of the above | Yes | 0 | 1 |
| 4 | 3.0 | Daily or almost daily | Yes, definitely | 17 | Man | No | Straight / Heterosexual | Indian | None of the above | No | 1 | 0 |
| 5 | 10.0 | Daily or almost daily | Neutral | 30 | Woman | Yes | Straight / Heterosexual | Asian | None of the above | No | 0 | 1 |
| 6 | 22.0 | Once a week | Neutral | 40 | Man | No | Bisexual | Black | None of the above | Yes | 1 | 0 |
| 7 | 5.0 | Multiple times per day | Yes, definitely | 30 | Man | No | Prefer to self-describe: | White | None of the above | Yes | 1 | 0 |

In [176...

```
df['MentalHealth'].replace({'I have a mood or emotional disorder (e.g., depression, bipolar disorder, etc.)': 'df
```

Out[176]:

| | YearsCodePro | BuyNewTool | SOPartFreq | Age | Gender | Trans | Sexuality | Ethnicity | Accessibility | MentalHealth | Data scientist or machine learning specialist | Developer, desktop or enterprise applications |
|---|--------------|------------------------|-----------------|-----|--------|-------|--------------------------|---------------|-------------------|---------------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| 0 | 14.0 | Multiple times per day | Neutral | 25 | Man | No | Bisexual | White | None of the above | I have a mood or emotional disorder (e.g., dep... | 1 | 0 |
| 1 | 20.0 | A few times per week | Yes, definitely | 35 | Man | No | Straight / Heterosexual | White | None of the above | No | 0 | 0 |
| 2 | 8.0 | Daily or almost daily | Yes, definitely | 25 | Man | No | Prefer to self-describe: | Not specified | None of the above | No | 0 | 0 |
| 3 | 15.0 | Multiple times per day | Yes, definitely | 25 | Man | No | Prefer to self-describe: | Not specified | None of the above | Yes | 0 | 1 |
| 4 | 3.0 | Daily or almost daily | Yes, definitely | 17 | Man | No | Straight / Heterosexual | Indian | None of the above | No | 1 | 0 |
| 5 | 10.0 | Daily or almost daily | Neutral | 30 | Woman | Yes | Straight / Heterosexual | Asian | None of the above | No | 0 | 1 |
| 6 | 22.0 | Once a week | Neutral | 40 | Man | No | Bisexual | Black | None of the above | Yes | 1 | 0 |
| 7 | 5.0 | Multiple times per day | Yes, definitely | 30 | Man | No | Prefer to self-describe: | White | None of the above | Yes | 1 | 0 |

In [180...

```
df = df.drop(index=0)
```

In [181...

```
df
```

Out[181]:

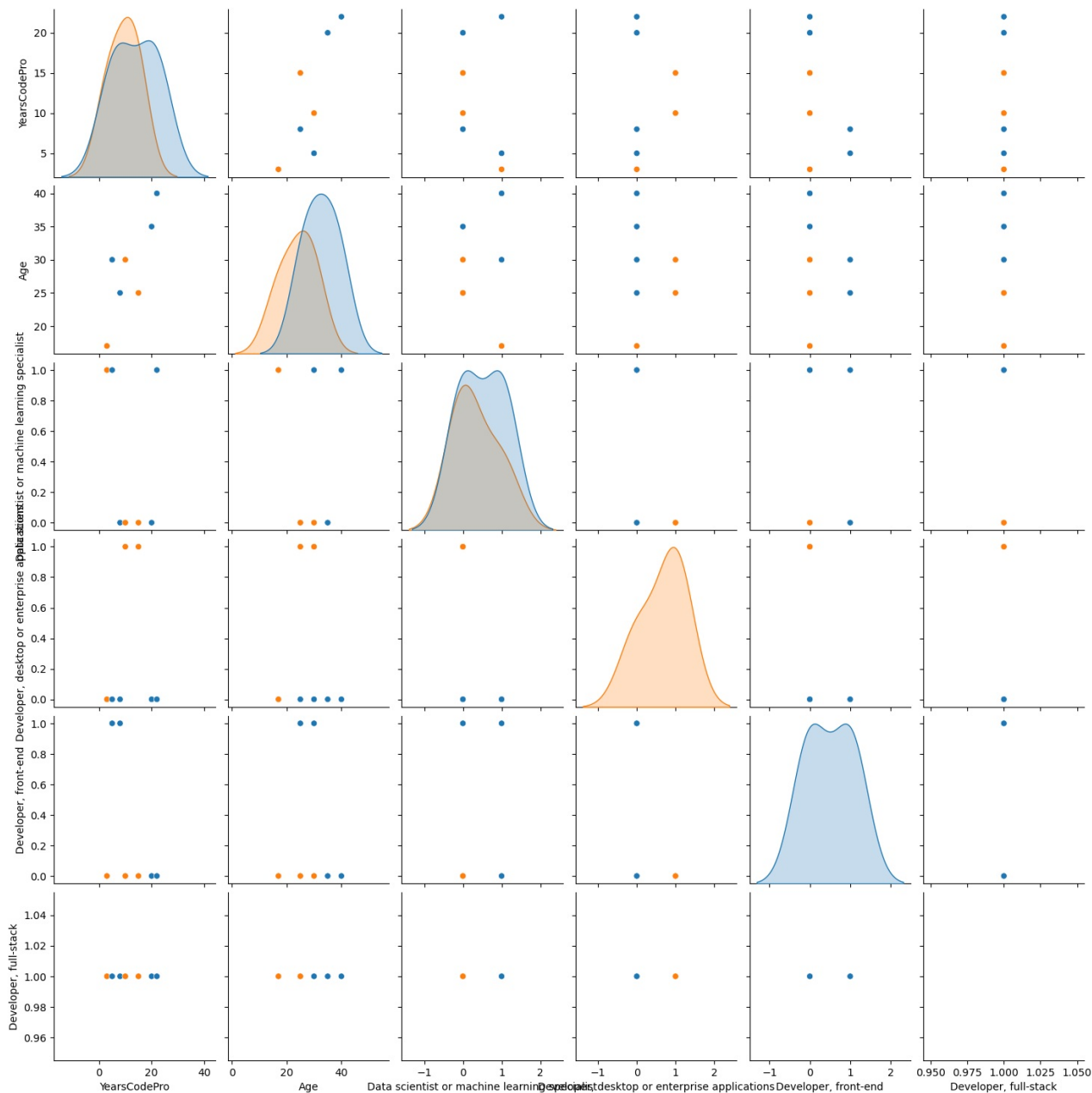
| | YearsCodePro | BuyNewTool | SOPartFreq | Age | Gender | Trans | Sexuality | Ethnicity | Accessibility | MentalHealth | Data scientist or machine learning specialist | Developer, desktop or enterprise applications |
|---|--------------|------------------------|-----------------|-----|--------|-------|--------------------------|---------------|-------------------|--------------|-----------------------------------------------|-----------------------------------------------|
| 1 | 20.0 | A few times per week | Yes, definitely | 35 | Man | No | Straight / Heterosexual | White | None of the above | No | 0 | 0 |
| 2 | 8.0 | Daily or almost daily | Yes, definitely | 25 | Man | No | Prefer to self-describe: | Not specified | None of the above | No | 0 | 0 |
| 3 | 15.0 | Multiple times per day | Yes, definitely | 25 | Man | No | Prefer to self-describe: | Not specified | None of the above | Yes | 0 | 1 |
| 4 | 3.0 | Daily or almost daily | Yes, definitely | 17 | Man | No | Straight / Heterosexual | Indian | None of the above | No | 1 | 0 |
| 5 | 10.0 | Daily or almost daily | Neutral | 30 | Woman | Yes | Straight / Heterosexual | Asian | None of the above | No | 0 | 1 |
| 6 | 22.0 | Once a week | Neutral | 40 | Man | No | Bisexual | Black | None of the above | Yes | 1 | 0 |
| 7 | 5.0 | Multiple times per day | Yes, definitely | 30 | Man | No | Prefer to self-describe: | White | None of the above | Yes | 1 | 0 |

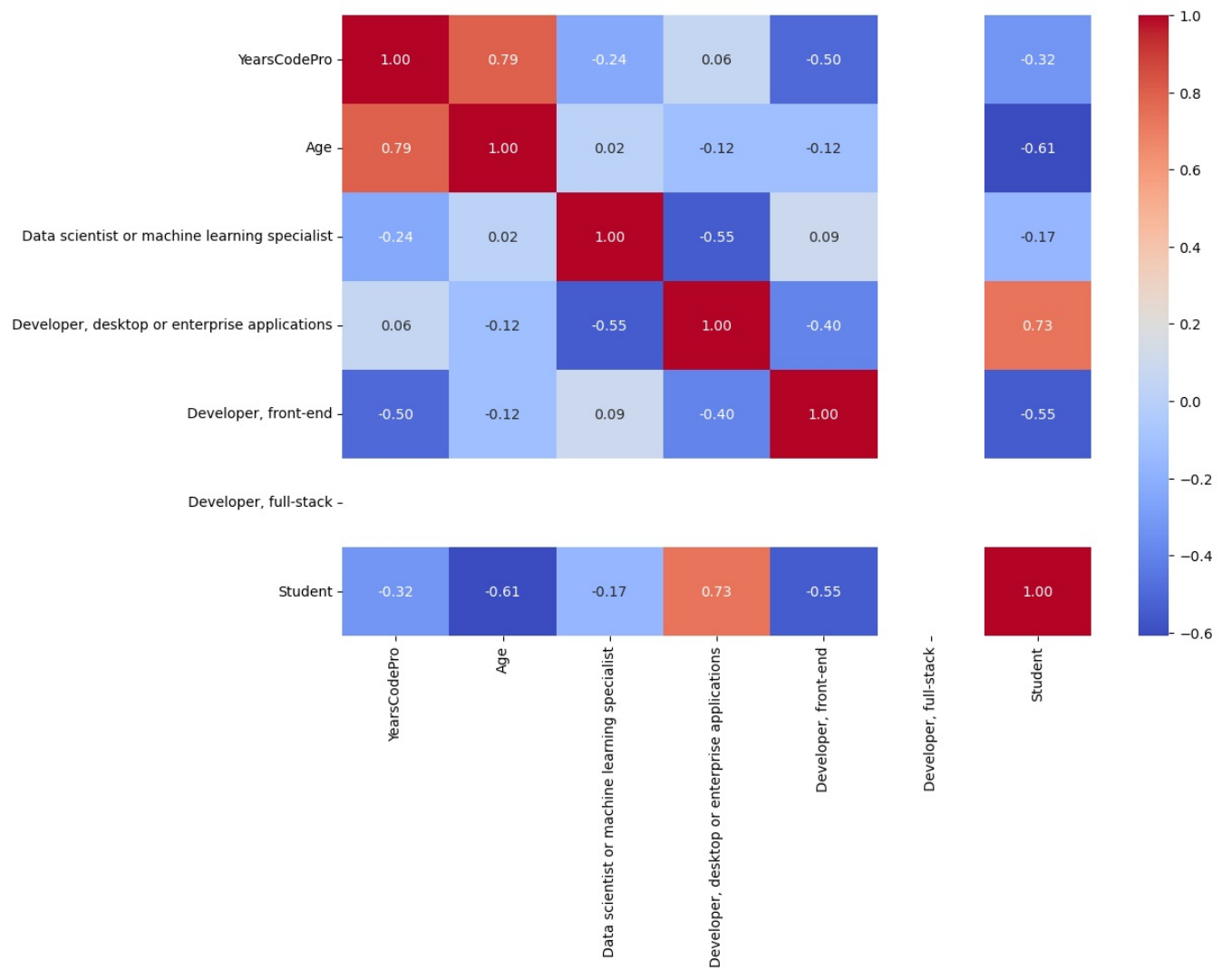
```
In [191]: import seaborn as sns
import matplotlib.pyplot as plt
# Pairplot to visualize relationships between numerical features
sns.pairplot(df_encoded[['YearsCodePro', 'Age', 'Data scientist or machine learning specialist', 'Developer, de
plt.show()

# Correlation heatmap
plt.figure(figsize=(12, 8))
correlation_matrix = df_encoded[['YearsCodePro', 'Age', 'Data scientist or machine learning specialist', 'Devel
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.show()

# Countplot for the target variable
plt.figure(figsize=(8, 6))
sns.countplot(x='TBranch', data=df_encoded)
plt.title('Distribution of Target Variable (TBranch)')
plt.show()
```

<Figure size 800x600 with 0 Axes>






```

-----
ValueError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_9700\185216870.py in <module>
    13 # Countplot for the target variable
    14 plt.figure(figsize=(8, 6))
--> 15 sns.countplot(x='TBranch', data=df_encoded)
    16 plt.title('Distribution of Target Variable (TBranch)')

    17 plt.show()

~\anaconda3\lib\site-packages\seaborn\_decorators.py in inner_f(*args, **kwargs)
    44     )
    45     kwargs.update({k: arg for k, arg in zip(sig.parameters, args)})
--> 46     return f(**kwargs)
    47     return inner_f
    48

~\anaconda3\lib\site-packages\seaborn\categorical.py in countplot(x, y, hue, data, order, hue_order, orient, color, palette, saturation, dodge, ax, **kwargs)
    3596     raise ValueError("Cannot pass values for both `x` and `y`")
    3597
-> 3598     plotter = _CountPlotter(
    3599         x, y, hue, data, order, hue_order,
    3600         estimator, ci, n_boot, units, seed,

~\anaconda3\lib\site-packages\seaborn\categorical.py in __init__(self, x, y, hue, data, order, hue_order, estimator, ci, n_boot, units, seed, orient, color, palette, saturation, errcolor, errwidth, capsize, dodge)
    1582         errwidth, capsize, dodge):
    1583         """Initialize the plotter."""
-> 1584         self.establish_variables(x, y, hue, data, orient,
    1585                                order, hue_order, units)
    1586         self.establish_colors(color, palette, saturation)

~\anaconda3\lib\site-packages\seaborn\categorical.py in establish_variables(self, x, y, hue, data, orient, order, hue_order, units)
    151         if isinstance(var, str):
    152             err = "Could not interpret input '{}".format(var)
--> 153             raise ValueError(err)
    154
    155         # Figure out the plotting orientation

ValueError: Could not interpret input 'TBranch'

```

In [183]: df.describe()

Out[183]:

| | YearsCodePro | Age | Data scientist or machine learning specialist | Developer, desktop or enterprise applications | Developer, front-end | Developer, full-stack | Other (please specify): | Student |
|-------|--------------|-----------|-----------------------------------------------|-----------------------------------------------|----------------------|-----------------------|-------------------------|----------|
| count | 7.000000 | 7.000000 | 7.000000 | 7.000000 | 7.000000 | 7.0 | 7.000000 | 7.000000 |
| mean | 11.857143 | 28.857143 | 0.428571 | 0.285714 | 0.285714 | 1.0 | 0.142857 | 0.428571 |
| std | 7.335498 | 7.470577 | 0.534522 | 0.487950 | 0.487950 | 0.0 | 0.377964 | 0.534522 |
| min | 3.000000 | 17.000000 | 0.000000 | 0.000000 | 0.000000 | 1.0 | 0.000000 | 0.000000 |
| 25% | 6.500000 | 25.000000 | 0.000000 | 0.000000 | 0.000000 | 1.0 | 0.000000 | 0.000000 |
| 50% | 10.000000 | 30.000000 | 0.000000 | 0.000000 | 0.000000 | 1.0 | 0.000000 | 0.000000 |
| 75% | 17.500000 | 32.500000 | 1.000000 | 0.500000 | 0.500000 | 1.0 | 0.000000 | 1.000000 |
| max | 22.000000 | 40.000000 | 1.000000 | 1.000000 | 1.000000 | 1.0 | 1.000000 | 1.000000 |

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []: