**CHAPTER 2**

**THE DATA**

**2.1 Introduction**

Data sourcing is key process of and data mining project because the whole project and outcomes is relying on this data. To get data for my project of finding trending technologies different data sources are available but I have to choose the one which is most reliable and trusted. Mostly tech people rely on web data sources, so I do the same. First, I have to choose the methodologies to get the data from web data sources.

**2.2 Methodologies to get Data.**

There are multiple methods available in data mining to get the data from data sources, but I am trying only one method to get data from web data source which are listed below:

* **API** (Data is collected through an Application Programing Interface)
* **Surveys** (Data is collected through trusted and verified surveys from well-known Organizations.)
* **Web Scraping** (Data is collect through scraping different websites.)

**2.3 Data Sources**

The authenticity and reliability of any data mining project is depending upon the authentic and reliable data source. Hence, I have to choose the data source carefully and after a lot of research I got to some authentic and most reliable data sources which are market leaders, and no one could question their authenticity. These data sources are listed as:

* **Stacks Overflow Developers Survey 2020**

Stacks Overflow is the leaders of the developer community and are used on daily basis from the tech professionals. I use stacks overflow survey as data source for my project.

**2.4 The Data**

The data from data sources is collected through different methodologies. Now the data is available for my project and the properties of this data is given below:

**The Developers Survey Data is collected from Stacks Overflow Survey so the questions of the survey along with response column is given below:**

| **Column Name** | **Question Text** |
| --- | --- |
| Respondent | Randomized respondent ID number (not in order of survey response time) |
| MainBranch | Which of the following options best describes you today? Here, by "developer" we mean "someone who writes code." |
| Hobbyist | Do you code as a hobby? |
| OpenSourcer | How often do you contribute to open source? |
| OpenSource | How do you feel about the quality of open source software (OSS)? |
| Employment | Which of the following best describes your current employment status? |
| Country | In which country do you currently reside? |
| Student | Are you currently enrolled in a formal, degree-granting college or university program? |
| EdLevel | Which of the following best describes the highest level of formal education that you have completed? |
| UndergradMajor | What was your main or most important field of study? |
| EduOther | Which of the following types of non-degree education have you used or participated in? Please select all that apply. |
| OrgSize | Approximately how many people are employed by the company or organization you work for? |
| DevType | Which of the following describe you? Please select all that apply. |
| YearsCode | Including any education, how many years have you been coding? |
| Age1stCode | At what age did you write your first line of code or program? (E.g., webpage, Hello World, Scratch project) |
| YearsCodePro | How many years have you coded professionally (as a part of your work)? |
| CareerSat | Overall, how satisfied are you with your career thus far? |
| JobSat | How satisfied are you with your current job? (If you work multiple jobs, answer for the one you spend the most hours on.) |
| MgrIdiot | How confident are you that your manager knows what they're doing? |
| MgrMoney | Do you believe that you need to be a manager to make more money? |
| MgrWant | Do you want to become a manager yourself in the future? |
| JobSeek | Which of the following best describes your current job-seeking status? |
| LastHireDate | When was the last time that you took a job with a new employer? |
| LastInt | In your most recent successful job interview (resulting in a job offer), you were asked to... (check all that apply) |
| FizzBuzz | Have you ever been asked to solve FizzBuzz in an interview? |
| JobFactors | Imagine that you are deciding between two job offers with the same compensation, benefits, and location. Of the following factors, which 3 are MOST important to you? |
| ResumeUpdate | Think back to the last time you updated your resumé, CV, or an online profile on a job site. What is the PRIMARY reason that you did so? |
| CurrencySymbol | Which currency do you use day-to-day? If your answer is complicated, please pick the one you are most comfortable estimating in. |
| CurrencyDesc | Which currency do you use day-to-day? If your answer is complicated, please pick the one you are most comfortable estimating in. |
| CompTotal | What is your current total compensation (salary, bonuses, and perks, before taxes and deductions), in Currency Symbol? Please enter a whole number in the box below, without any punctuation. If you are paid hourly, please estimate an equivalent weekly, monthly, or yearly salary. If you prefer not to answer, please leave the box empty. |
| CompFreq | Is that compensation weekly, monthly, or yearly? |
| ConvertedComp | Salary converted to annual USD salaries using the exchange rate on 2019-02-01, assuming 12 working months and 50 working weeks. |
| WorkWeekHrs | On average, how many hours per week do you work? |
| WorkPlan | How structured or planned is your work? |
| WorkChallenge | Of these options, what are your greatest challenges to productivity as a developer? Select up to 3: |
| WorkRemote | How often do you work remotely? |
| WorkLoc | Where would you prefer to work? |
| ImpSyn | For the specific work you do, and the years of experience you have, how do you rate your own level of competence? |
| CodeRev | Do you review code as part of your work? |
| CodeRevHrs | On average, how many hours per week do you spend on code review? |
| UnitTests | Does your company regularly employ unit tests in the development of their products? |
| PurchaseHow | How does your company make decisions about purchasing new technology (cloud, AI, IoT, databases)? |
| PurchaseWhat | What level of influence do you, personally, have over new technology purchases at your organization? |
| LanguageWorkedWith | Which of the following programming, scripting, and markup languages have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the language and want to continue to do so, please check both boxes in that row.) |
| LanguageDesireNextYear | Which of the following programming, scripting, and markup languages have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the language and want to continue to do so, please check both boxes in that row.) |
| DatabaseWorkedWith | Which of the following database environments have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the database and want to continue to do so, please check both boxes in that row.) |
| DatabaseDesireNextYear | Which of the following database environments have you done extensive development work in over the past year and which do you want to work in over the next year? (If you both worked with the database and want to continue to do so, please check both boxes in that row.) |
| PlatformWorkedWith | Which of the following platforms have you done extensive development work for over the past year? (If you both developed for the platform and want to continue to do so, please check both boxes in that row.) |
| PlatformDesireNextYear | Which of the following platforms have you done extensive development work for over the past year? (If you both developed for the platform and want to continue to do so, please check both boxes in that row.) |
| WebFrameWorkedWith | Which of the following web frameworks have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the framework and want to continue to do so, please check both boxes in that row.) |
| WebFrameDesireNextYear | Which of the following web frameworks have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the framework and want to continue to do so, please check both boxes in that row.) |
| MiscTechWorkedWith | Which of the following other frameworks, libraries, and tools have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the technology and want to continue to do so, please check both boxes in that row.) |
| MiscTechDesireNextYear | Which of the following other frameworks, libraries, and tools have you done extensive development work in over the past year, and which do you want to work in over the next year? (If you both worked with the technology and want to continue to do so, please check both boxes in that row.) |
| DevEnviron | Which development environment(s) do you use regularly? Please check all that apply. |
| OpSys | What is the primary operating system in which you work? |
| Containers | How do you use containers (Docker, Open Container Initiative (OCI), etc.)? |
| BlockchainOrg | How is your organization thinking about or implementing blockchain technology? |
| BlockchainIs | Blockchain / cryptocurrency technology is primarily: |
| BetterLife | Do you think people born today will have a better life than their parents? |
| ITperson | Are you the "IT support person" for your family? |
| OffOn | Have you tried turning it off and on again? |
| SocialMedia | What social media site do you use the most? |
| Extraversion | Do you prefer online chat or IRL conversations? |
| ScreenName | What do you call it? |
| SOVisit1st | To the best of your memory, when did you first visit Stack Overflow? |
| SOVisitFreq | How frequently would you say you visit Stack Overflow? |
| SOVisitTo | I visit Stack Overflow to... (check all that apply) |
| SOFindAnswer | On average, how many times a week do you find (and use) an answer on Stack Overflow? |
| SOTimeSaved | Think back to the last time you solved a coding problem using Stack Overflow, as well as the last time you solved a problem using a different resource. Which was faster? |
| SOHowMuchTime | About how much time did you save? If you're not sure, please use your best estimate. |
| SOAccount | Do you have a Stack Overflow account? |
| SOPartFreq | How frequently would you say you participate in Q&A on Stack Overflow? By participate we mean ask, answer, vote for, or comment on questions. |
| SOJobs | Have you ever used or visited Stack Overflow Jobs? |
| EntTeams | Have you ever used Stack Overflow for Enterprise or Stack Overflow for Teams? |
| SOComm | Do you consider yourself a member of the Stack Overflow community? |
| WelcomeChange | Compared to last year, how welcome do you feel on Stack Overflow? |
| SONewContent | Would you like to see any of the following on Stack Overflow? Check all that apply. |
| Age | What is your age (in years)? If you prefer not to answer, you may leave this question blank. |
| Gender | Which of the following do you currently identify as? Please select all that apply. If you prefer not to answer, you may leave this question blank. |
| Trans | Do you identify as transgender? |
| Sexuality | Which of the following do you currently identify as? Please select all that apply. If you prefer not to answer, you may leave this question blank. |
| Ethnicity | Which of the following do you identify as? Please check all that apply. If you prefer not to answer, you may leave this question blank. |
| Dependents | Do you have any dependents (e.g., children, elders, or others) that you care for? |
| SurveyLength | How do you feel about the length of the survey this year? |
| SurveyEase | How easy or difficult was this survey to complete? |

**Table 1.1 Survey Result column and question text**

**Source: Stacksoverflow Developers Survey. [**https://insights.stackoverflow.com/survey**]**

**Here is the sample of Dataset:**

Table

Description automatically generated

**Table 1.2 Sample of Dataset**

**Number of columns and rows:**

Number of rows : 64461

Number of columns : 61

**Statistical Summary of Data:**

**Table

Description automatically generated**

**Table 1.3 Statistical Summary**

**Data Types:**

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Respondent 14281 non-null int64

1 Age 13768 non-null float64

2 ConvertedComp 13620 non-null float64

3 CompFreq 14281 non-null object

4 CompTotal 14281 non-null float64

5 Country 14281 non-null object

6 DatabaseDesireNextYear 14281 non-null object

7 DatabaseWorkedWith 14281 non-null object

8 DevType 14281 non-null object

9 EdLevel 14281 non-null object

10 Gender 14281 non-null object

11 LanguageDesireNextYear 14281 non-null object

12 LanguageWorkedWith 14281 non-null object

13 MiscTechDesireNextYear 14281 non-null object

14 MiscTechWorkedWith 14281 non-null object

15 OpSys 14281 non-null object

16 PlatformDesireNextYear 14281 non-null object

17 PlatformWorkedWith 14281 non-null object

18 UndergradMajor 14281 non-null object

19 WebframeDesireNextYear 14281 non-null object

20 WebframeWorkedWith 14281 non-null object

21 WorkWeekHrs 14281 non-null float64

22 YearsCodePro 14281 non-null object

23 NormalizedAnnualCompensation 14281 non-null float64

dtypes: float64(5), int64(1), object(18)

**Table 1.4 Data Types**

For Complete Summary of the Dataset please download the html report generated from

Pandas profiling from the link below:

Link: https://github.com/jonygujjar/Final-Year-Project/blob/5db05dc13ead1f445847e7fd6588f29ebba054f2/survey\_report.html