# **Dataset exploration and analysis**

**Question 2**

1. Describe the data in the data set:

* What is the data?

The dataset created is based on people who filled a survey, and it tells if the person currently works at a consulting company; if yes, then which consulting firm it is. What is their current title, location, the compensation in USD, bonus, and how many hours do they work in a week? Moreover, it asks if they received an offer from any other consulting company this year. If yes, then the company's name, location, offered compensation, bonus, and if they accepted the offer or not.

* What is the industry or area of knowledge of the?

The data is related to the industry of consulting jobs.

* What is the range of records included?

The records range is all those who landed on Consulting humor's survey on Instagram and filled the survey. Therefore, our dataset includes people who filled the survey, which has details about their current employment and the details regarding the future offer from any company if applicable.

* Is the data structured/unstructured?

Structured; on observing the data, we can identify that all the fields have data that have number/categorize/date or binary values.

1. Data Source

* What were the sources of this data?

The dataset selected is from Kaggle. The data was collected through survey filling of the 'Consulting Humor' Instagram community. This survey was posted on the community's Instagram stories, and people interested and willing to fill the survey filled it.

* How do you think the data was collected?

The data collected was based on people’s answers in the survey. The data is based on their current consulting job statistics and on the job offer they have in hand from another consulting company.

1. Describe each field (column) of the dataset

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Column # | Name of column | Description | Type | Format | Valid Range | Actual Range |
| A | Timestamp | This column provides information regarding the date and time when the user filled the survey. | Structured | Quantitative -Continuous | For the time, the survey was available | From 9/29/2021 10:35:35 to 10/19/2021 10:58:20 |
| B | Do you CURRENTLY work at a consulting firm? | They were asking the user if they currently worked at a consulting firm. | Structured | Qualitative - Boolean | Yes/No | Yes/No |
| C | Which firm do you CURRENTLY work for? | If the user answers 'yes' in #2, they will fill in the name of their current consulting company. Otherwise, they cannot answer this question. | Structured | Qualitative – Categories | All the consulting firms in the world. | Only those unique consulting firms whose employees have filled the survey. |
| D | What is your CURRENT title (or equivalent)? | If the user answered 'yes' in #2, they would fill the job position that the user is currently at in the present consulting company. | Structured | Qualitative – Categories | All the titles from interns, analyst/staff till partner level. | This column includes the current titles of employees who have filled the survey. |
| E | What COUNTRY are you CURRENTLY based out of? | What is the country's name where the user currently has their job? | Structured | Qualitative – Categories | All the countries in the world. | The countries from which people have filled the survey. |
| F | What is your CURRENT annual base compensation in USD (not including bonuses, perks, or other incentives)? Please consult Google for conversion tools for local currency into USD. Enter number values only. | The user needs to enter the annual base compensation that the company is paying them in US dollars. This column will accept numerical values. | Structured | Quantitative – Discrete | $1-$n | $1 to $1630000, $0 for those who didn’t want to share their compensation. |
| G | What is the total amount in USD of BONUSES you estimate you will receive in 2021? Enter a number only. | The user needs to enter the estimated bonus amount given to them in US dollars. This column will accept numerical values. | Structured | Quantitative – Discrete | $1-$n | $1 - $50000000  $0 is for those who did not want to share their bonus or did not receive any. |
| H | Expected total compensation (calculated) | The total compensation in US dollars is calculated in this column by adding the annual base compensation and the bonus amount. | Structured | Quantitative – Discrete | $1-$n | $2 - $50000001  $0 is for those who did not want to share their compensation. |
| I | Have you gotten an OFFER from another company this year? | This column tells if the user has received a job offer from any other company in the current year. | Structured | Qualitative – Boolean | Yes/No | Yes/No |
| J | Was the offer from another CONSULTING FIRM? | The user could only answer this question if they answered 'yes' in #9. Furthermore, this asks if the offer received was from a consulting company. | Structured | Qualitative – Boolean | Yes/No | Yes/No |
| K | Which firm was the HIGHEST (base compensation) OFFER from? | The user could only answer this question if the user answered 'yes' in #10. If the user received offers from multiple consulting companies, then which company offered the highest annual base compensation. | Structured | Qualitative – Categories | All the consulting firms in the world. | Only those consulting firms from which people have filled the survey. |
| L | What was the OFFERED title (or equivalent)? | What was the job position offered by the new consulting company? | Structured | Qualitative – Categories | All the titles from interns, analyst/staff till partner level. | This column includes the titles of the people who have filled the survey. |
| M | What COUNTRY was the OFFER based out of? | This column provides information regarding the new job location. | Structured | Qualitative – Categories | All the countries in the world. | This column includes locations of the future company of the people who filled the survey. |
| N | What was the HIGHEST OFFERED annual base compensation in USD (not including bonuses, perks, or other incentives)? Please consult Google for conversion tools for local currency into USD. Enter a number only. | This column asks about the new consulting company's annual base compensation in US dollars. | Structured | Quantitative – Discrete | $1 -$n | $1 - $ 1300000.  $0 and blanks for those who didn't want to share the compensation. |
| O | What was the total annual BONUS OFFERED? | This column asks about the new consulting company's annual bonus in US dollars. | Structured | Quantitative – Discrete | $1 -$n | $1 - $6900000  $0 and blanks for those who didn't want to share the bonus. |
| P | Did you accept the offer? | This column tells if the user accepted/declined or is thinking about the offer currently. | Structured | Qualitative – Categories | Yes, no, still thinking, unsure, waiting for other offers. | Yes, no, and still thinking. |

**Question 4**

Descriptive Questions

1. Which is the consulting company that is currently hiring?

* We can use the column ‘Which firm was the HIGHEST (base compensation) OFFER from?'
* to check from which companies the respondents got an offer.
* We can know that this company was/is hiring by knowing this information. After which, people looking for a job change can also apply to these companies for an opportunity.

1. People from which consulting company have the highest compensation?

* We can use the columns ‘Which firm do you CURRENTLY work for?’ and ‘Expected total compensation (calculated).'
* Using these columns, we can order the salaries in descending order and then filter on the company names to know which company is paying how much. Finally, we can identify the companies that are providing the best offers.
* People who want to switch jobs for better compensation can apply to these companies using this information.

**Descriptive Question**

1. Why are people switching their jobs?

* We can use the column ‘Have you gotten an OFFER from another company this year?' and filter only for 'yes' values to extract this information.
* Combining the filtered column with the ‘Which firm do you CURRENTLY work for?’ and ‘Expected total compensation (calculated).' We can analyze if the people who wanted to switch jobs had low compensation.

**Predictive Question**

1. Will, more people with low compensation will switch their jobs?

* We can use the columns ‘Have you gotten an OFFER from another company this year?' and 'Expected total compensation (calculated)' to know the compensation of people who have received an offer from another company.
* We can analyze the current compensation of people who have another offer by knowing this. If the compensation is low, we can predict that more people will apply for jobs elsewhere.
* The response variable would be ‘future company switch rate’.