## Appendix A: Samples of Datasets Used in Empirical Study

**Figure 6 Snippet of Dataset 1.**

| | sl_no | gender | ssc_p | ssc_b | hsc_p | hsc_b | hsc_s | degree_p | degree_t | workex | etest_p | specialisation | mba_p | status | salary |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | M | 67.00 | Others | 91.00 | Others | Commerce | 58.00 | Sci&Tech | No | 55.0 | Mkt&HR | 58.80 | Placed | 270000.0 |
| 1 | 2 | M | 79.33 | Central | 78.33 | Others | Science | 77.48 | Sci&Tech | Yes | 86.5 | Mkt&Fin | 66.28 | Placed | 200000.0 |
| 2 | 3 | M | 65.00 | Central | 68.00 | Central | Arts | 64.00 | Comm&Mgmt | No | 75.0 | Mkt&Fin | 57.80 | Placed | 250000.0 |
| 3 | 4 | M | 56.00 | Central | 52.00 | Central | Science | 52.00 | Sci&Tech | No | 66.0 | Mkt&HR | 59.43 | Not Placed | NaN |
| 4 | 5 | M | 85.80 | Central | 73.60 | Central | Commerce | 73.30 | Comm&Mgmt | No | 96.8 | Mkt&Fin | 55.50 | Placed | 425000.0 |

**Figure 7 Snippet of Dataset 2.**

| | enrollee_id | city | city_development_index | gender | relevent_experience | enrolled_university | education_levevel | major_discipline | experience | company_size | company_type | last_new_job | training_hours | target |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 8949 | city_103 | 0.920 | Male | Has relevent experience | no_enrollment | Graduate | STEM | >20 | NaN | NaN | 1 | 36 | 1.0 |
| 1 | 29725 | city_40 | 0.776 | Male | No relevent experience | no_enrollment | Graduate | STEM | 15 | 50-99 | Pvt Ltd | >4 | 47 | 0.0 |
| 2 | 11561 | city_21 | 0.624 | NaN | No relevent experience | Full time course | Graduate | STEM | 5 | NaN | NaN | never | 83 | 0.0 |
| 3 | 33241 | city_115 | 0.789 | NaN | No relevent experience | NaN | Graduate | Business Degree | <1 | NaN | Pvt Ltd | never | 52 | 1.0 |
| 4 | 666 | city_162 | 0.767 | Male | Has relevent experience | no_enrollment | Masters | STEM | >20 | 50-99 | Funded Startup | 4 | 8 | 0.0 |

## Appendix B: Datasets 1 and 2: Detailed Description

**Table 9 Dataset 1**

| Column Name | Data Type | Range of Values | Missing Values (%) |
|---|---|---|---|
| sl_no | int64 | 1 to 215 | 0 |
| gender | object | M, F | 0 |
| ssc_p | float64 | 40.89 to 89.4 | 0 |
| ssc_b | object | Central, Others | 0 |
| hsc_p | float64 | 37 to 97.7 | 0 |
| hsc_b | object | Central, Others | 0 |
| hsc_s | object | Commerce, Science, Arts | 0 |
| degree_p | float64 | 50 to 91 | 0 |
| degree_t | object | Comm&Mgmt, Sci&Tech, Others | 0 |
| workex | object | Yes, No | 0 |
| etest_p | float64 | 50 to 98 | 0 |
| specialisation | object | Mkt&Fin, Mkt&HR | 0 |
| mba_p | float64 | 51.21 to 77.89 | 0 |
| status | object | Placed, Not Placed | 0 |
| salary | float64 | 20,000 to 94,000 | 31.2 |

**Table 10 Dataset 2**

| Column Name | Data Type | Range of Values | Missing Values (%) |
|---|---|---|---|
| enrollee_id | int64 | 1 to 33380 | 0 |
| city | object | City_1, City_2, ... City_123 | 0 |
| city_development_index | float64 | 0.448 to 0.949 | 0 |
| gender | object | Male, Female, Other | 23.5 |
| relevant_experience | object | No relevant experience, Has relevant experience | 0 |
| enrolled_university | object | no_enrollment, Part time course, Full time course | 2 |
| education_level | object | Primary School, High School, Graduate, Masters, Phd | 2.4 |
| major_discipline | object | No Major, Arts, Business Degree, Humanities, STEM, Other | 14.7 |
| experience | object | <1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, >20 | 0.3% |
| company_size | object | <10, 10/49, 50-99, 100-500, 500-999, 1000-4999, 5000-9999, 10000+ | 31 |
| company_type | object | Pvt Ltd, Funded Startup, Public Sector, Early Stage Startup, NGO, Other | 32 |
| last_new_job | object | never, 1, 2, 3, 4, >4 | 2.2 |
| training_hours | int64 | 1 to 336 | 0 |
| target | float64 | 0, 1 | 0 |

## Appendix C: List of Datasets and Data Wrangling Recipes

**Table 11 Dataset 1**

Source Link: https://www.kaggle.com/datasets/benroshan/factors-affecting-campus-placement.

Date Accessed: 3rd February 2023

| No. | Link |
|---|---|
| 1 | https://www.kaggle.com/code/benroshan/you-re-hired-analysis-on-campus-recruitment-data |
| 2 | https://www.kaggle.com/code/priyankarao18/eda-visualization |
| 3 | https://www.kaggle.com/code/arindambaruah/placement-predictions-using-log-reg-knn-rfc-xgb |
| 4 | https://www.kaggle.com/code/avaniudupa/campusrecruitment-predict-if-a-student-gets-placed |
| 5 | https://www.kaggle.com/code/achintyatripathi/highcharts-eda-logistic-reg-prediction |
| 6 | https://www.kaggle.com/code/adityaaggarwal09/a-bull-s-eye-on-prediction |
| 7 | https://www.kaggle.com/code/poojaravi01/campus-placement-prediction-model |
| 8 | https://www.kaggle.com/code/ami4501/kernel3fdc106f66 |
| 9 | https://www.kaggle.com/code/mani97/placed-or-not-eda-classification-88-8 |
| 10 | https://www.kaggle.com/code/saitej31/campus-recruitment-90-acc-logistic-regression |

**Table 12 Dataset 2**

Source Link:

https://www.kaggle.com/datasets/arashnic/hr-analytics-job-change-of-data-scientists.

Date Accessed: 3rd February 2023

| No. | Link |
|---|---|
| 1 | https://www.kaggle.com/code/josephchan524/hranalytics-lightgbm-classifier-auc-80 |
| 2 | https://www.kaggle.com/code/gokulrajkmv/hr-analytics-attrition-prediction |
| 3 | https://www.kaggle.com/code/elyousfiomar/hyperparameter-tuning-gradient-boosting |
| 4 | https://www.kaggle.com/code/gsdeepakkumar/looking-for-a-job |
| 5 | https://www.kaggle.com/code/stuartday274/job-change-predictions-using-a-pipeline |
| 6 | https://www.kaggle.com/code/ihormuliar/hr-analytics |
| 7 | https://www.kaggle.com/code/italomarcelo/predictions-hr-job |
| 8 | https://www.kaggle.com/code/anmol686/job-change-prediction |
| 9 | https://www.kaggle.com/code/harshbhatnagar/hr-analystics-prediction |
| 10 | https://www.kaggle.com/code/nicklcorona/predict-data-science-job-change |
| 11 | https://www.kaggle.com/code/scharom/old-but-interpretable-the-logit-model |
| 12 | https://www.kaggle.com/code/gcmadhan/hr-analysis-data-scientist-job |
| 13 | https://www.kaggle.com/code/raj5kumar5/hr-analytics-job-change |
| 14 | https://www.kaggle.com/code/artyomkolas/hr-analytics |
| 15 | https://www.kaggle.com/code/meet1415/beginner-notebook-for-hr-analytics |
| 16 | https://www.kaggle.com/code/ouyangwenguang/hr-analytics-job-change-of-data-scientists |
| 17 | https://www.kaggle.com/code/naiara/hr-analytics-job |
| 18 | https://www.kaggle.com/code/adrianabellanoza/aabellanoza-hr-analytics-job-change |
| 19 | https://www.kaggle.com/code/kfhuang/hr-analytics |
| 20 | https://www.kaggle.com/code/kowalskyyy999/lightgbm-predict-hr-analytics |
| 21 | https://www.kaggle.com/code/bashtovoy/notebooke3cc6c23a8 |
| 22 | https://www.kaggle.com/code/navesammy/hr-analytics |
| 23 | https://www.kaggle.com/code/nicolamignoni/job-change-of-data-scientists |