# Data Transformation Plan

## Applicants CSVs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Raw data** | **ERD** | | **Transformation Required** | **Missing Data** |
| **Column** | **Data Type** |
| id |  |  |  |  |
| name | applicant.first\_name  applicant.last\_name | str | Split at first space  Title case | Drop record |
| gender | applicant.gender | str | Title case  Male, female, non-binary, prefer not to disclose | “Prefer not to disclose” |
| dob | applicant.date\_of\_birth | datetime | UK format  Remove time value | Placeholder |
| email | applicant.email | str | Lower case Validate: @ | Placeholder |
| city | post\_code.city | str | Title case | “Unknown” |
| address | address.address\_line | str | Title case | “Unknown” |
| postcode | post\_code.post\_code | str | Capital case | “Unknown” |
| phone\_number |  |  | Drop |  |
| University | applicant.university | str | Title Case | “Unknown”/  “Did Not Attend” |
| degree | applicant.university\_grade | str | Remove both leading 0s -> 02:01 becomes 2:1 | “None” |
| invited\_date | sparta\_day\_results.invited\_date | datetime | Combine with month | Ridiculous placeholder |
| month | sparta\_day\_results.invited\_date | datetime | Combine with invited\_date | Ridiculous placeholder |
| invited\_by | sparta\_day\_results.invited\_by | str | Title case | “Not invited” |

## Course Behaviour Results CSVs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Raw data** | **ERD** | | **Transformation Required** | **Missing Data** |
| **Column** | **Data Type** |
| trainer | course.trainer\_name | str | Title case | Fill with name in other records |
| file title  e.g.  Data\_29\_2019\_03\_04 | course.subject\_name  course.start\_date  course.class\_number | str  datetime  int | Title Case  UK format, remove time value |  |
| column headers  e.g  Analytic\_W1 | behaviour\_scores.week | int | Extract num from string |  |
| Analytic\_W(n) | behaviour\_scores.analytical | int | Match column heading number to correct week value | 0 |
| Independent\_W(n) | behaviour\_scores.independent | int | Match column heading number to correct week value | 0 |
| Determined\_W(n) | behaviour\_scores.determined | int | Match column heading number to correct week value | 0 |
| Professional\_W(n) | behaviour\_scores.professional | int | Match column heading number to correct week value | 0 |
| Studious\_W(n) | behaviour\_scores.studious | int | Match column heading number to correct week value | 0 |
| Imaginative\_W(n) | behaviour\_scores.imaginative | int | Match column heading number to correct week value | 0 |

## Sparta Day .txt files

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Raw data** | **ERD** | | **Transformation Required** | **Missing Data** |
| **Column** | **Data Type** |
| Date | sparta\_days.date | datetime | UK format |  |
| Academy location | sparta\_days.location | str | Title Case |  |
| Psychometric score | sparta\_day\_results.psychometrics | int | Remove “/100”  Ensure int |  |
| Presentation score | sparta\_day\_results.presentation | int | Remove “/32”  Ensure int |  |

## Applicant Results JSONs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Raw data** | **ERD** | | **Transformation Required** | **Missing Data** |
| **Column** | **Data Type** |
| tech\_self\_score  { <tech\_name> : <score> } | sparta\_day\_results\_tech\_self\_score.tech\_name  sparta\_day\_results\_tech\_self\_score.score | str  int | Each dictionary entry is new record | Don’t add |
| strengths :  [… <strengths>] | sparta\_day\_results\_strengths.name | str | Each list item is new record | Don’t add |
| weaknesses :  [… < weaknesses >] | sparta\_day\_results\_ weaknesses.name | str | Each list item is new record | Don’t add |
| Self\_development | sparta\_day\_results.self\_development | bool |  | False |
| Geo\_flex | sparta\_day\_results.geo\_flex | bool |  | False |
| financial\_support\_self | sparta\_day\_results. financial\_support\_self | bool |  | True |
| result | sparta\_day\_results. result | bool |  | False |
| course\_interest | sparta\_day\_results.course\_interest | str | Title Case | “No interest” |