

Course Project

Tools for Decision Making 2022-01

Context and Justification

One of the most common tasks in engineering management is to analyse vast quantities of data to understand what's happening and taking optimal decisions based on our findings.

Description

Groups of 3 students will develop the project.

Laura, the new CEO of *The Institute of Equality, Diversity and Inclusiveness* in region Corellia has contacted us to discuss what is happening in two cities (Alderaan and Bespin). After the hard work of data collection, the new CEO has been able to collect information about the selection processes over two months. This information contains data on the candidates and their environment, the job position, the recruiting company and its environment, and whether the candidate finally received a call back on his or her application.

Laura is concerned about several issues. She believes that certain companies may be biased towards some candidates because of their gender, race, or even background they come from. Laura has therefore asked us to carry out an exploratory analysis.

Based on this information, we need to conduct an exploratory data analysis and prepare a technical report (no page limit, but meaningful), a presentation with a maximum of 10 slides (including title and final slides) and four high-quality graphs.

Note: It is highly recommended to seek external sources of information (either in dataset or report formats) for the analysis and the reporting.

Deliverables

The project has two deliverables:

- Each group must upload a single file with the technical report in pdf/html/MS Word format to the platform ATENEA.
- Each group must upload a single file with the presentation in pdf/html/MS PowerPoint format to the platform ATENEA.

The name of the files will be: '2022-PROJ01-inGroup.pdf'.

Assessment

The grade of the activity depends on three elements:

- The number of tools and techniques that we have learnt in the course.
- The level of difficulty (or depth) of the analysis.
- The accuracy and quality of the description of the analysis.
- The structure and the order of the report.

Data Set Information

A data set with 4870 observations of 57 variables.

Information from the candidates' resume

- education - Highest education, with levels of 0 = not reported; 1 = high school diploma; 2 = high school graduate; 3 = some college; 4 = college or more.
- n_jobs - Number of jobs listed on resume.
- years_exp - Number of years of work experience on the resume.
- honors - Indicator variable for which 1 = resume mentions some honors.
- volunteer - Indicator variable for which 1 = resume mentions some volunteering experience.
- military - Indicator variable for which 1 = resume mentions some military experience.
- occup_broad - Occupation broad with levels 1 = executives and managerial occupations, 2 = administrative supervisors, 3 = sales representatives, 4 = sales workers, 5 = secretaries and legal assistants, 6 = clerical occupations
- work_in_school - Indicator variable for which 1 = resume mentions some work experience while at school
- email - Indicator variable for which 1 = email address on applicant's resume.
- computer_skills - Indicator variable for which 1 = resume mentions some computer skills.
- special_skills - Indicator variable for which 1 = resume mentions some special skills.
- first_name - Applicant's first name.
- sex - Sex, with levels of 'f' = female; 'm' = male.
- race - Race, with levels of 'b' = black; 'w' = white.
- h - Indicator variable for which 1 = high-quality resume.
- l - Indicator variable for which 1 = low-quality resume.
- col - Indicator variable for which 1 = applicant has a college degree or more.

Information regarding if the candidate was pre-selected

- call - Indicator variable for which 1 = applicant was called back.

Information regarding the candidates' context and environment

- `frac_black` - Fraction of blacks in applicant's zip.
- `frac_white` - Fraction of whites in applicant's zip.
- `l_med_hh_inc` - Log median household income in applicant's zip.
- `frac_dropout` - Fraction of high-school dropouts in applicant's zip.
- `frac_colp` - Fraction of college degree or more in applicant's zip
- `l_inc` - Log per capita income in applicant's zip.

Information regarding the job position

- `city` - City, with levels of 'c' = city_a; 'b' = city_b.
- `kind` - Kind, with levels of 'a' = administrative; 's' = sales.
- `ad_id` - Employment ad identifier.
- `expminreq` - Minimum experience required, if any (in years when numeric).
- `school_req` - Specific education requirement, if any. 'hsg' = high school graduate, 'somcol' = some college, 'colp' = four year degree or higher
- `manager` - Indicator variable for which 1 = executives or managers wanted.
- `supervisor` - Indicator variable for which 1 = administrative supervisors wanted.
- `secretary` - Indicator variable for which 1 = secretaries or legal assistants wanted.
- `sales_rep` - Indicator variable for which 1 = sales representative wanted.
- `retail_sales` - Indicator variable for which 1 = retail sales worker wanted.
- `req` - Indicator variable for which 1 = ad mentions any requirement for job.
- `exp_req` - Indicator variable for which 1 = ad mentions some experience requirement.
- `com_req` - Indicator variable for which 1 = ad mentions some communication skills requirement.
- `educ_req` - Indicator variable for which 1 = ad mentions some educational requirement.
- `comp_req` - Indicator variable for which 1 = ad mentions some computer skill requirement.
- `org_req` - Indicator variable for which 1 = ad mentions some organizational skills requirement.

Information regarding the employer

- `manuf` - Indicator variable for which 1 = employer industry is manufacturing.

- trans_com - Indicator variable for which 1 = employer industry is transport or communication.
- bank_real - Indicator variable for which 1 = employer industry is finance, insurance or real estate.
- trade - Indicator variable for which 1 = employer industry is wholesale or retail trade.
- bus_service - Indicator variable for which 1 = employer industry is business or personal services.
- oth_service - Indicator variable for which 1 = employer industry is health, education or social services.
- ownership - Ownership status of employer, with levels of 'non-profit'; 'private'; 'public'
- parent_sales - Sales of parent company (in millions of US \$).
- parent_emp - Number of parent company employees.
- branch_sales - Sales of branch (in millions of US \$).
- branch_emp - Number of branch employees.

Information regarding the employers' context and environment

- frac_black_emp_zip - Fraction of blacks in employers's zipcode.
- frac_white_emp_zip - Fraction of whites in employer's zipcode.
- l_med_hh_inc_emp_zip - Log median household income in employer's zipcode.
- frac_dropout_emp_zip - Fraction of high-school dropouts in employer's zipcode.
- frac_colp_emp_zip - Fraction of college degree or more in employer's zipcode.
- l_inc_emp_zip - Log per capita income in employer's zipcode.