

Factors Affecting Job Satisfaction:

Database Documentation

The primary database in this Job Satisfaction project contains 531,216 rows of data that was pushed to one SQL database table, which was obtained from the IPUMS Higher Ed site (<https://higher.ed.ipums.org/>). The data available there was gathered through the following surveys:

National Survey of College Graduates 2003, SESTAT
Survey of Doctorate Recipients 2003, SESTAT

The data used for this project covers surveys conducted in 2003, 2006, 2008, 2010, and 2013. The following factors obtained from IPUMS were used for analysis.

<u>COLUMN (FACTOR)</u>	<u>DATA TYPE</u>
AGE	int64
GENDER	object
RACETH	object
CHTOT	int64
DGRDG	object
SALARY	int64
JOBSATIS	int64

From this original database of 531,216 rows, any NA values (and their equivalents) were excluded from analysis. The data was then queried using SQL to aggregate by factor type to obtain the average job satisfaction rate for each factor by each factor category. The results were saved to individual dataframes for each factor, and then those smaller dataframes were consolidated into one main dataframe containing the following columns.

<u>COLUMN</u>	<u>DATA TYPE</u>
factor_cat	object
avg_job_satisfaction	float64
factor	object
factor_id	object

Data from this main dataframe above is fed to the front-end dashboard using JupyterDash, which connects to an external interface where the user is able to select the desired factor to see how it affects job satisfaction. The dashboard is available at <http://127.0.0.1:8050/>.