Factors Affecting Job Satisfaction: Database Documentation

The primary database in this Job Satisfaction project contains 531,216 rows of data, which was obtained from the IPUMS Higher Ed site (https://highered.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 were obtained from IPUMS for analysis.

COLUMN (FACTOR)	DATA TYPE
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 (or their equivalents) were excluded from analysis. The data was then queried using SQL to aggregate by factor type to obtain the mean job satisfaction rate for each factor by each factor category. The resulting dataframe contains the following columns.

COLUMN	DATA TYPE
factor_cat	object
avg_job_satisfaction	float64
factor	object
factor_id (pk)	object

This data was then fed to a 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/.

The documentation should list the name of every feature in each database, its data type, whether the feature is a primary key or a foreign key that links to a primary key in another table, and short notes that describe what the feature means. The documentation must provide accurate information about every feature, and should describe all of the data that is present in every database.

Ease of use (5 points): is the design of the documentation intuitive, and does it include enough information about the databases to help a user navigate them?

ER diagram (5 points): if at least one database is relational, is there a complete and accurate ER diagram to accompany the documentation?