Emily Wong
BAX 421
Section 1

Database Creation

Dataset Conversion

To convert my dataset, I first imported the entire CSV dataset into one table in MySQL using the table import tool. From there, I cleaned the dataset to remove rows with no response (data). Once initial non-responses were removed, I proceeded to create the tables that I required, and selected columns from the loaded full dataset to easily create the tables I needed. Afterwards, I created a 'respondent_id' column to differentiate each row/response in the dataset. Once the column was created, I added it as a primary key to the 'respondent' table.

Since my dataset contains census responses, each respondent has 0 or 1 answer to a question (column). So, the 'respondent_id' in the respondent table is a foreign key for every other table in the dataset. I created each foreign key and referenced 'respondent_id'. Once the keys were added to each of the tables, I moved on to removing empty rows in each table. Once that was done, I double-checked each table to ensure that the keys, rows, and columns were accurate and roughly matched my ERD. My tables differ slightly, as I realized there were more ways to sort and organize my main tables.

Challenges

As the questions in the census in which the data was collected are not mandatory, it was difficult to determine a non-respondent versus a respondent that sparsely answered the census questions when cleaning the dataset for proper conversion. I narrowed non-responses down to be considered rows that contained less than 4 responses.

Another challenge when importing the data was deriving the primary and foreign keys for my tables. MySQL doesn't allow for the creation of a primary or foreign key when using SELECT, which was required for me to create my tables. So, I had to create my tables, then manually add in the keys to each table. This proved very difficult, as I kept getting NULL values for my references when adding foreign keys. However, I was able to figure this out after experimenting with column creations using ALTER and ADD, or using the table tool.

Data Dictionary

Overview of database (taken from SQL)

Name	Engine	Version	Row Format	Rows
creative_income	InnoDB	10	Dynamic	1109
creative_spend	InnoDB	10	Dynamic	720
experience	InnoDB	10	Dynamic	2009
industry	InnoDB	10	Dynamic	933
presenter	InnoDB	10	Dynamic	215
presenter_pressures	InnoDB	10	Dynamic	175
presenter_talent	InnoDB	10	Dynamic	161
respondent	InnoDB	10	Dynamic	2058

creative_income

Column Name	Type	Description
Respondent_id (FK)	int	Foreign key reference to 'respondent' table, identifies census respondent

income_performance_touring	int	Annual income from touring
income_recordings	int	Annual income from recording music
income_songwriting	int	Annual income from songwriting
income_studio_work	int	Annual income from studio work
income_merchandise	int	Annual income from merchandise sales
income_teaching	int	Annual income from teaching
paid_performances_per_month	int	Number of paid performances per month
paid_gigs_base_guarantee	int	Number of gigs that are paid with base guarantee

creative_spend

Column Name	Type	Description
Respondent_id (FK)	int	Foreign key reference to ‘respondent’ table, identifies census respondent
annual_spend_recordings	int	Annual spend on recordings
annual_spend_promotion	int	Annual spend on promotions
annual_spend_social_media	int	Annual spend on social media promotions
annual_spend_supplies	int	Annual spend on supplies
annual_spend_workspace	int	Annual spend on workspace rentals
annual_spend_gear_rentals	int	Annual spend on gear rentals
annual_spend_merchandise	int	Annual spend on merchandise
annual_spend_legal	int	Annual spend on legal fees, accounting

experience

Column Name	Type	Description
Respondent_id (FK)	int	Foreign key reference to ‘respondent’ table, identifies census respondent
music_sector	text	Respondent role within the music sector
years_experience	int	Years of music experience
music_education	int	Highest level of music education (if applicable)
training_topics	text	Suggested training topics for entering musicians

training_level	text	Suggested training level for entering musicians
intent_continue_career_3yrs	text	Intent to continue music for the next 3 years (T/F)
community_business_participation	text	Level of involvement in communities, list of engagements
work_performance_space_status	text	Status of respondent's access to work or performance spaces

industry

Column Name	Type	Description
Respondent_id (FK)	int	Foreign key reference to 'respondent' table, identifies census respondent
professional_experience	text	Industry worker's types professional experience in the industry
work_related_to_music_percentage	text	Percentage of work related to music production
austin_client_ranking	text	Ranking of proportion of clients that are Austin locals
usa_client_ranking	text	Ranking of proportion of clients that are based in the US
international_client_ranking	text	Ranking of proportion of clients that are international

presenter

Column Name	Type	Description
Respondent_id (FK)	int	Foreign key reference to 'respondent' table, identifies census respondent
ownership_structure	text	Ownership structure (for-profit, non-profit) of presenter venue
venue_type	text	Type of music venue
role_live_music	text	When live music performances occur in the venue
venue_capacity	int	Capacity of the venue

presenter_pressures

Column Name	Type	Description
Respondent_id (FK)	int	Foreign key reference to 'respondent' table, identifies census respondent
talent_costs	int	Ranking 1-10 of how much pressure a presenter faces regarding talent costs
crowd_work	int	Ranking 1-10 of how much pressure a presenter faces regarding crowd work
labor	int	Ranking 1-10 of how much pressure a presenter faces regarding labor
property_tax	int	Ranking 1-10 of how much pressure a presenter faces regarding property tax

marketing	int	Ranking 1-10 of how much pressure a presenter faces regarding marketing
permits	int	Ranking 1-10 of how much pressure a presenter faces regarding permits
unpredicted_costs	int	Ranking 1-10 of how much pressure a presenter faces regarding unpredicted costs
building_operations	int	Ranking 1-10 of how much pressure a presenter faces regarding building operation costs

presenter_talent

Column Name	Type	Description
Respondent_id (FK)	int	Foreign key reference to ‘respondent’ table, identifies census respondent
guarantee_as_business_expense	int	Percentage of talent paid through guarantees as a business expense
talent_paid_door_proceeds	int	Percentage of talent paid through a portion of door proceeds
talent_paid_door_proceeds_only	int	Percentage of talent paid through door proceeds only
talent_paid_fixed_bar_sales	int	Percentage of talent paid through fixed bar sales
talent_paid_tips_only	int	Percentage of talent paid through tips only
local_performers_ratio	int	Ratio of local performers to non-local performers

respondent

Column Name	Type	Description
Respondent_id (PK)	int	Primary key that identifies census respondent
county	text	Respondent county
distanceDT	int	Respondent distance from their downtown area
race	text	Respondent race
hispanic	text	Respondent race (hispanic or not), T/F
age	int	Respondent age
gender	text	Respondent gender
sexual_orientation	text	Respondent sexual orientation

Business Questions

To reiterate, my dataset is a music census data for the Greater Austin area in Texas. The following are some of the business questions that I want to use SQL queries to potentially answer.

1. How profitable is the music scene in the Greater Austin area for music creatives?
2. What is the average age of respondents based on their level of music education?
3. What is the average age of respondents based on their experience in the music industry?
4. How many creatives have a profitable music career?
5. What proportion of respondents lack a work or music space?
6. What is the race demographic of respondents?
7. What is the average years of experience a respondent has based on their race? Based on gender and sexuality?
8. What is the distribution of respondents based on their role in the music scene (creative, presenter, industry)?
9. What aspect of their jobs do presenters face the most pressure?