Married at First Sight Statistical Analysis

Introduction: The goal of the project was to analyze the reality show "Married at First Sight" (MAFS) to determine if the experiment of matching individuals to marry without prior acquaintance is a success or failure. The show is currently running on its 17th season, but has enough damage been done already?

Data Description: To collect the data, kaggle.com was checked to see if existing MAFS data was already gathered. This was done as I have not learned Python, I do not have the ability to do web scraping at this time. The original dataset was scraped from by roliepoly. Wikipedia was used to bring the dataset up to date through the 16th season. The data is in a .csv file with a size of 5 kb. The dataset can be found at https://www.kaggle.com/datasets/caseyoztel/married-at-first-sight-up-to-season-16.
The entire dataset in Excel has 26 columns and 128 rows, including header. The columns are named: id, couple_id, season, premier_date, month_year_release, location, name, age, age_bracket, age_difference, gender, job, job_type, decision, spilt_decision, status, and then data for the various experts.

Research Objectives: Analyze the reality show "Married at First Sight" (MAFS) to determine if the experiment of matching individuals who marry without prior acquaintance is a success or failure. My three null hypotheses were (1) marriage success is equal across job types, (2) marriage success is equal across locations, and (3) marriage success on the show is not equal to the US average.

Methodology:
Exploratory Analysis: Many summary statistics were found. The initial decision percentage was calculated. The age of contestant minimum, maximum, mode, and standard deviation, along with the largest and smallest age difference. Job type to successful marriage was broken down. Many various pivot tables were created to get to know the dataset.

Hypothesis (1) Marriage success is equal across job types.
A frequency table was generated with the job types compared to the number of successful and unsuccessful marriages. Based on the divorce rate, an expected frequency was calculated. After calculating the expected frequency, a chi-squared test was run and compared to the set p-value to reject at 0.05.

Hypothesis (2) Marriage success is equal across locations.
A frequency table was created with the various job locations compared to the marriage success. The expected frequency for both successful and unsuccessful marriage was calculated. The chi-squared test was run and compared to the set p-value to reject at 0.05.

Hypothesis (3) Marriage success on the show is not equal to the US average.
A frequency table was generated showing each year of the show and the divorce percent. Another frequency table was created using US Census data for all states in the United States divorce rates per year. A line chart was generated using this data to visually understand the data points. The average divorce percent for the whole term (2014-2021) was calculated. Next, setting the p-value to reject the null hypothesis at 0.05, the chi-squared test was run.