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.

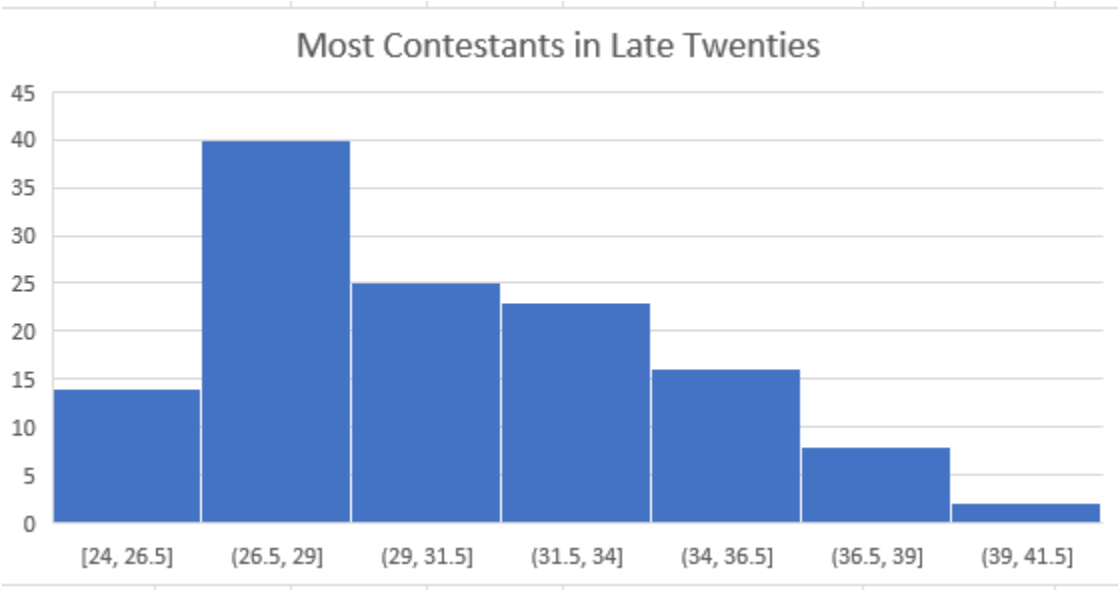
Hypothesis (1) Marriage success is equal across job types.
A frequency table was generated with the job types compared to the number of successful and unsuccessful marriages. Based on the divorce rate, an expected frequency was calculated. After calculating the expected frequency, a chi-squared test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

Data:

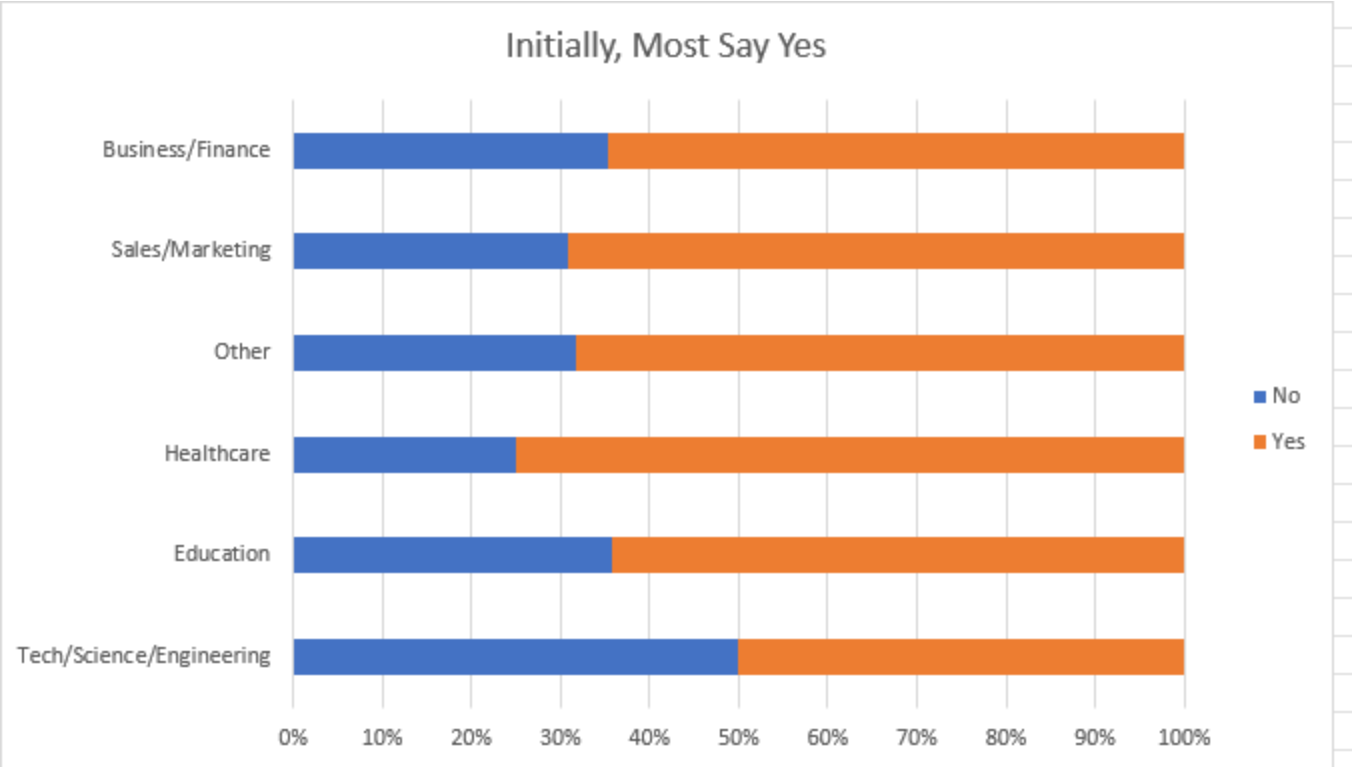
Age MIN	Age MAX	Age MODE	Age Standard Deviation	Age Difference MIN	Age Difference MAX
24	41	29	3.64	0	10

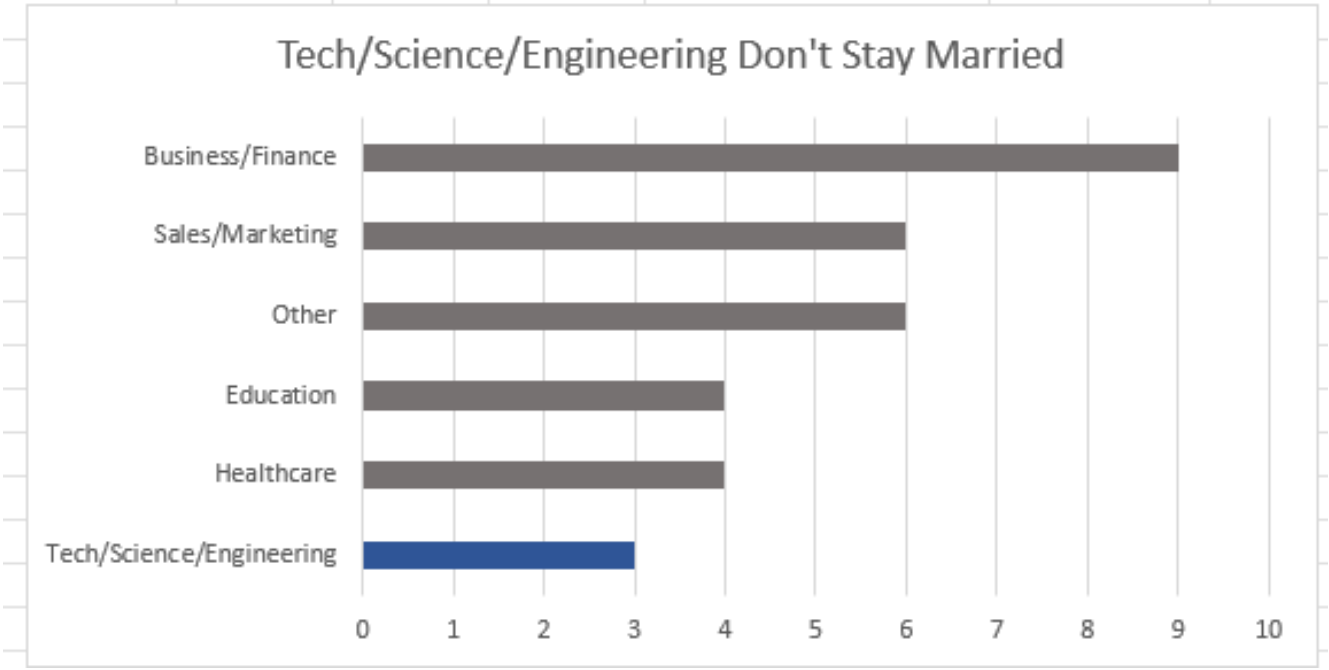


Hypothesis (1) Marriage success is equal across job types.

Job Type	Divorced	Married		Total	Expected Frequency Divorced	chi sqrd test
Business/Finance	25	9		34	18.75	0.1562356
Education	10	4		14	7.5	
Healthcare	12	4		16	9	
Other	16	6		22	12	
Sales/Marketing	20	6		26	15	
Tech/Science/Engineering	13	3		16	9.75	
	96	32				

H0	Marriage success IS EQUAL across job types	FAIL TO REJECT
H1	Marriage success IS NOT EQUAL across job types	

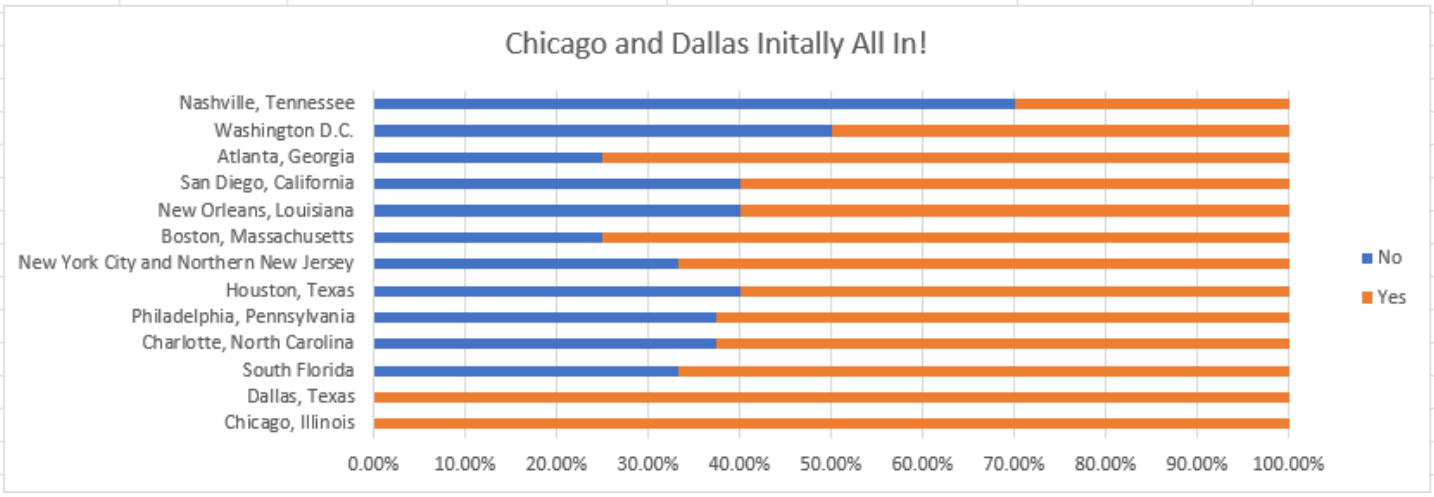




Hypothesis (2) Marriage success is equal across locations.

Location	Unsuccessful Marriage	Successful Marriage	Total Marriage	Expected Frequency Unsuccessful	Expected Frequency Successful	CHI SQRD TEST UNSUC	degrees of freedom	CHI SQRD STAT
New Orleans, Louisiana	4	6	10	3	1.5	0.785130387	13	0.0088086
Atlanta, Georgia	10	6	16	7.5	1.5			0.3060656
Charlotte, North Carolina	4	4	8	3	1			0.0088086
Philadelphia, Pennsylvania	4	4	8	3	1			0.0088086
Washington D.C.	8	2	10	6	0.5			0.1563997
New York City and Northern New Jersey	10	2	12	7.5	0.5			0.3060656
Boston, Massachusetts	14	2	16	10.5	0.5			0.626156
Chicago, Illinois	4	2	6	3	0.5			0.0088086
Dallas, Texas	4	2	6	3	0.5			0.0088086
Nashville, Tennessee	8	2	10	6	0.5			0.1563997
San Diego, California	10	0	10	7.5	0			0.3060656
Houston, Texas	10	0	10	7.5	0			0.3060656
South Florida	6	0	6	4.5	0			0.053847
TOTAL	96	32	128					

H0	Marriage success IS EQUAL across locations.	FAIL TO REJECT NULL
H1	Marriage success IS NOT EQUAL across locations.	



Hypothesis (3) Marriage success on the show is not equal to the US average.

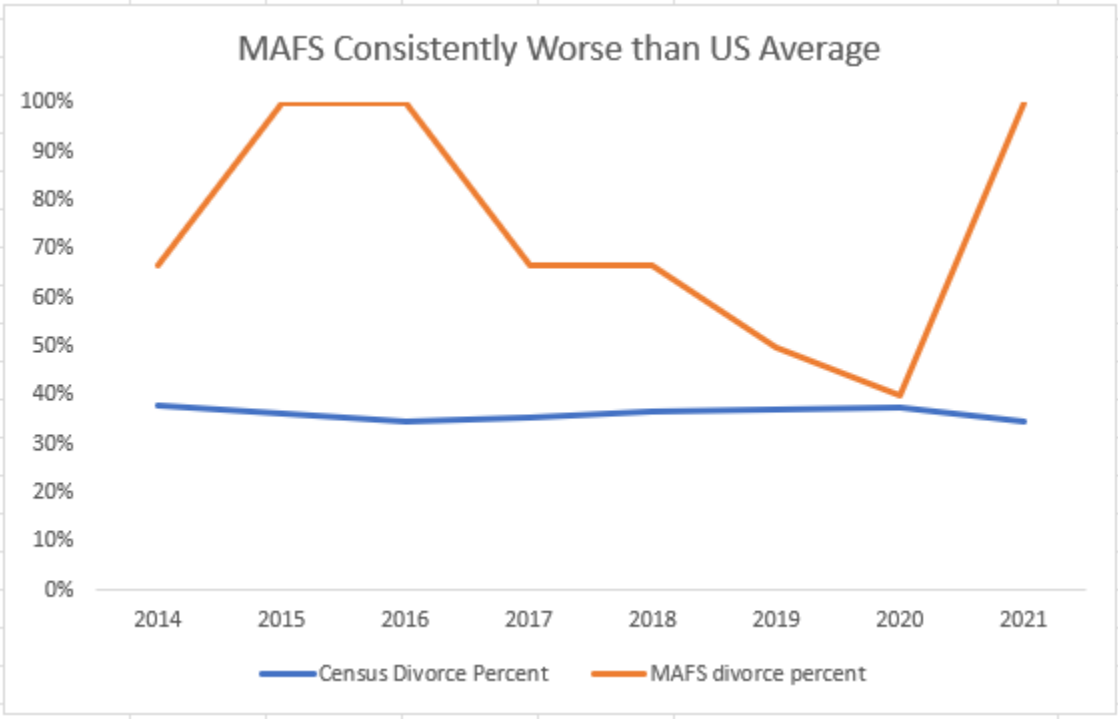
US CENSUS DATA

Year	Marriages	Divorces	Divorce Percent
2014	2,140,272	813,862	38%
2015	2,221,579	800,909	36%
2016	2,251,411	776,288	34%
2017	2,236,496	787,251	35%
2018	2,132,853	782,038	37%
2019	2,015,603	746,971	37%
2020	1,676,911	630,505	38%
2021	1,985,072	689,308	35%

MAFS Season Breakdown for Divorce

year	season	Marriages	Divorces	Divorce_percent
2014	1	3	2	67%
2015	2	3	3	100%
2015	3	3	3	100%
2016	4	3	3	100%
2017	5	3	2	67%
2018	6	3	2	67%
2018	7	3	2	67%
2019	8	4	2	50%
2019	9	4	2	50%
2020	10	5	4	80%
2020	11	5	2	40%
2021	12	5	2	40%
2021	13	5	5	100%
2022	14	5	5	100%
2022	15	5	5	100%
2023	16	5	4	80%

Year	Census Divorce Percent	MAFS divorce percent
2014	38%	67%
2015	36%	100%
2016	34%	100%
2017	35%	67%
2018	37%	67%
2019	37%	50%
2020	38%	40%
2021	35%	100%



Census Avg	36%		p-value to reject	0.05
MAFS AVG	75%			
CHISQRTEST	0.7332498			REJECT NULL HYPOTHESIS
H0	Marriage success on the show is EQUAL to US average			
H1	Marriage success on the show is NOT EQUAL to US average			

Discussion:

The analysis came to two conclusions in my mind. (1) The creators of the show did the best they could to keep things fair across the locations and types of contestants they brought on. They picked locations that didn't show bias, and same with job types of contestants which I imagine feeds into the personality of the contestants. (2) The show is not a success for the contestants if they truly are coming onto the show to find their one and only spouse.

While finalizing my statistical analysis, I realized that my analysis and my hypothesis didn't quite match. My analysis used the numbers for divorce, while the hypothesis was written about marriage success. It may not matter, but I will reflect on the purpose of precise hypothesis moving forward.

Conclusion:
The show is not a success in terms of relationships being long lasting, but I imagine that it is a success to the creators and advertisers, as the show is going onto its 17th season.

I feel that it is important for the contestants going onto this show to understand this data so they understand this is likely not leading to a lasting love, just a short stint on TV. I can imagine if I was truly trying to find love, I would stay very far away from the "experts" on the show. Although, further analysis needs to be done regarding the success of each expert.