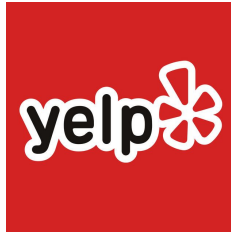# Order Up

By: Mindy Zhou & Ngoc Tran
July 16, 2019

Introduction

- Start a restaurant business
  - Location
  - Cuisine
  - Price Range
  - Hours of Operation
- Data
  - Yelp Business Details Data
    - 6728 observations
  - Metrics
    - review_count
    - rating

Location

One-way-anova, Welch-anova and Tukey-HSD:

- H_0: DC, MD, VA are the same
- H_A: DC, MD, VA are not the same

review_count:

```
    Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====================================================================
group1 group2  meandiff  p-adj   lower       upper      reject
---------------------------------------------------------------------
   DC     MD  -113.4217  0.001  -113.7438  -113.0996    True
   DC     VA   -39.1391  0.001   -39.4613   -38.817     True
   MD     VA    74.2826  0.001    73.9604    74.6047    True
---------------------------------------------------------------------
```

rating:

```
    Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====================================================================
group1 group2 meandiff  p-adj   lower     upper    reject
---------------------------------------------------------------------
   DC     MD   -0.2888  0.001  -0.2897   -0.2879    True
   DC     VA   -0.0615  0.001  -0.0624   -0.0606    True
   MD     VA    0.2273  0.001   0.2264    0.2282    True
---------------------------------------------------------------------
```

Maryland Wins!!!

Cuisine

Modern American

Bar

t-test and Welch's t-test:

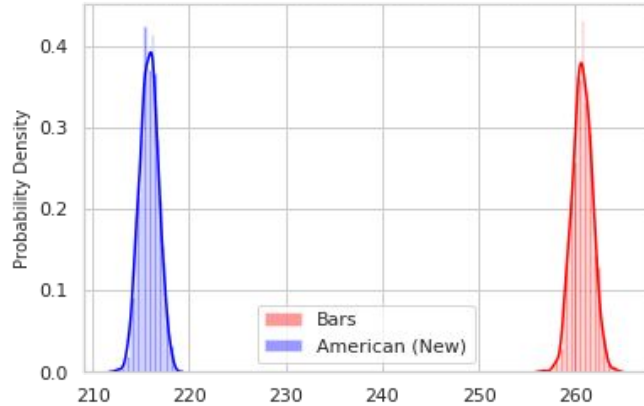- H_0: Bar is the same as American (New)
- H_1: Bar is not the same as American (New)
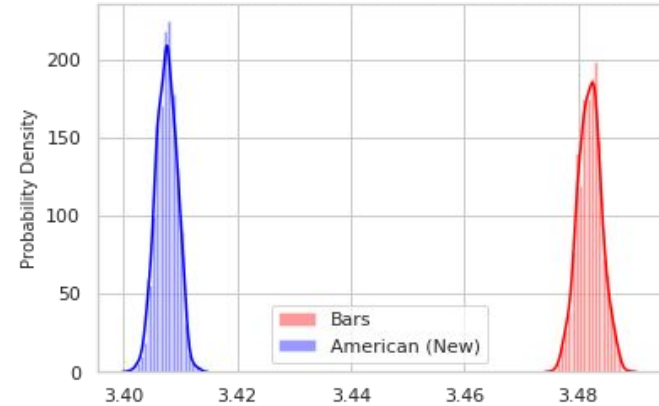

Probability Density Plot for Bars and American (New) - review_count - MD

Ttest_indResult(statistic=1011.6349050663866, pvalue=0.0)

Probability Density Plot for Bars and American (New) - rating - MD
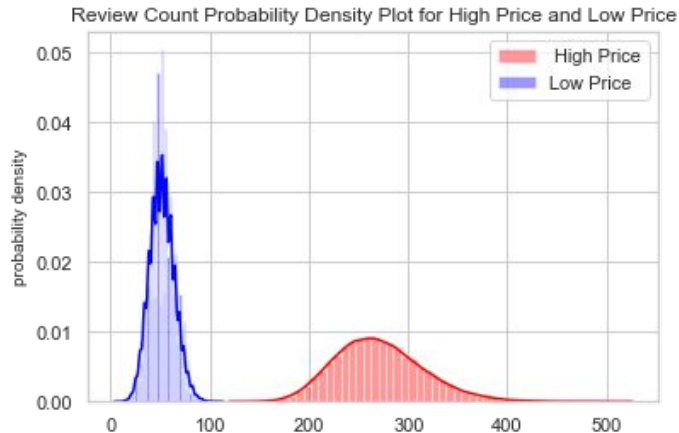
Ttest_indResult(statistic=853.6946379107173, pvalue=0.0)

Price Range

Bars

Welch's t-tests:

- H_0: > \$\$ is the same as <= \$\$
- H_1: > \$\$ is not the same as <= \$\$

Review Count Probability Density Plot for High Price and Low Price

Ttest_indResult(statistic=1498.0623445595186, pvalue=0.0)

Rating Probability Density Plot for High Price and Low Price

Ttest_indResult(statistic=945.5058504512784, pvalue=0.0)

Hours of Operation

Bar
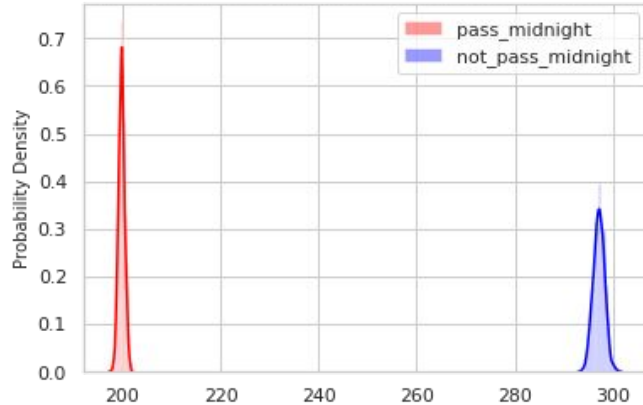
Welch's t-tests:

- H_0: pass_midnight is the same as not_pass_midnight
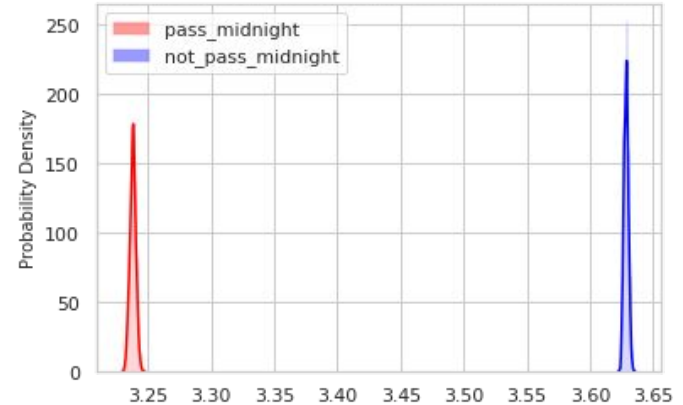- H_1: pass_midnight is not the same as not_pass_midnight

Probability Density Plot for pass_midnight and not_pass_midnight - review_count - MD

Ttest_indResult(statistic=-2383.4886945376957, pvalue=0.0)

Probability Density Plot for pass_midnight and not_pass_midnight - rating - MD

Ttest_indResult(statistic=-4237.331040257049, pvalue=0.0)

Conclusion

- Start a restaurant business
  - Location:
    - Maryland part of the DMV Area
  - Cuisine:
    - Bars
  - Price Range:
    - $$$ - $$$$
  - Hours of Operation:
    - Close before midnight

Questions?

THANK YOU!

Sources

https://www.yelp.com/developers/documentation/v3/business

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html

https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.levene.html

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.shapiro.html

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.normaltest.html

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.f_oneway.html

https://pingouin-stats.org/generated/pingouin.welch_anova.html#pingouin.welch_anova

https://docs.scipy.org/doc/scipy/reference/stats.html

https://www.statsmodels.org/stable/generated/statsmodels.stats.multicomp.pairwise_tukeyhsd.html