Please note that we are merely presenting one example outline of the summary. Feel free to deviate from this outline or not use this outline. Also, your summary must be in paragraph form, with reasonable section/subsection headings.

**Introduction/Motivation**: (a brief introduction about motivation/what the report is about/what you plan to describe)

1. Clearly lay out what domain/restaurants/category/area that you are planning to analyze
2. Clearly lay out specific questions you want to address in the report
   1. Example 1: Our analysis focuses on establishments that primarily serve tea and coffee shops in Madison and have been on Yelp from BLANK to BLANK. Among these establishments, our specific goals are to (list out questions/objectives/).
3. If needed, provide a short thesis statement about what you have found in your analysis.

**Background Information/Data Cleaning/Data Pre-Processing**:

1. Discuss any relevant/key aspects about the data (e.g. where did data come from, what format, etc.)
2. Discuss how you merged/cleaned/pre-processed data
   1. Example 1: To study reviewer’s sentiment, we combined BLANK and BLANK. We used software BLANK and BLANK provided by BLANK (cite reference). Using Yelp’s API guidelines, we filtered reviews with keywords BLANKL and BLANK. We also followed standard practice in NLP (cite references) and did BLANK steps to pre-process text data.
3. Discuss any outliers removed/fixed/imputed **even before building a model** (e.g. reviews with BLANK are BLANK, some predictors look suspicious, etc.)
   1. Example 1: I removed reviews that meet criterions BLANK because of BLANK.
   2. Example 2: I fixed/imputed reviews/predictors that meet criterions BLANK by BLANK because of BLANK
4. Discuss any new predictors/features that you may have created and why you created them.
   1. Example 1: We create new word predictors BLANK by using BLANK. We used the software BLANK to parse/combine words. We also followed suggestions from works by BLANK and BLANK (add references at the end) and added/removed predictors BLANK.

**Exploratory Data Analysis (EDA)**

1. Discuss any useful plots/tables/statistics that help the reader different aspects of the data. This sections should be designed in such a way to help you discuss (a) the business plan and (b) analytic insights.
   1. Example 1: Plot showing correlation between ratings and characteristics of review (e.g. length of review, word characters, etc.) or correlation between words
   2. Example 2: Table showing frequency/proportion of words by rating of reviews (1,2,3,4,5)
   3. Example 3: Averages (SEs) of mean ratings and/or distribution of ratings.
   4. Example 4: We find that some “neutral” words (“and”, “the”, etc.) are not correlated with ratings (corr = BLANK)
2. Interpret your EDA (in *laymen’s terms*\*\*)
   1. Example 1: Figure/Table 1 shows that the frequency of “good” and “excellent” in reviews are highly correlated with ratings (corr = 0.80).
   2. Example 2: Mean ratings are BLANK (SE = BLANK), with mean length of reviews (by word count) is BLANK (SE=BLANK)
3. (If needed, may overlap with Part 1) Concisely state your initial findings about the reviews
   1. Example 1: Our EDA finds that reviews in our dataset skews positive. Also, as expected, positive words are positive ratings and negative words are negatively correlated with negative ratings.
4. NOTE: Depending on the models that you use, this section may overlap a lot with Part 1 (or 2) below. If so, I would make sure to avoid redundancies and/or integrate the sections if you feel like it is appropriate.

**Part 1: Key Findings About Businesses BLANK/Areas BLANK OR Analysis of BLANK Market on Yelp**

1. Concisely and clearly state key analytical insights you found. Again, this may overlap with the EDA section and if so I would try to avoid any redundancies.
   1. Example 1: We find that on average, having a parking lot increases ratings by BLANK stars (SE:BLANK, pval:BLANK,)
   2. See other examples from previous lectures.
2. Discuss any important statistical tests you have conducted to support your observations/findings/EDA above (i.e. to show that what we observed is not due to random chance). This may be integrated with the first bullet numbered point to make your report more concise.
   1. Example 1: We fit a multiple linear regression model with outcome as ratings and predictors as BLANK (either state all predictors or group predictors into categories or provide some example predictors). The linear regression model has an R^2 of BLANK, implying a good fit.
   2. Example 2: Using the regression model in example 1, we conducted the following test to see whether having word BLANK we found is significant in predicting ratings, while controlling for other words. (briefly state hypothesis; which test statistic you used; p-value/test stat value; state the Type I error you’re willing to tolerate). From our statistical test/p-value, we can conclude that having word BLANK in the review leads to (interpret your statistical tests/p-values in a laymen’s term; see lecture notes for details).
   3. Example 3: The estimated slope and intercept are BLANK and BLANK, with 95% CIs BLANK and BLANK. This implies that BLANK (interpret in laymen’s terms). Also, based on the 95% CI, we can reject/retain the null hypothesis of BLANK and BLANK. In other words (interpret them in laymen’s terms)
3. Discuss any model diagnostics that you have done to support that the inference you done above (testing, CIs, p-values, etc.) is valid
   1. Include any model diagnostics you did. Include relevant plots/tables/etc.
   2. Briefly explain your model diagnostics (e.g. why you did it, what assumption is this diagnostic checking, what is your conclusion from the diagnostic checks, how did you resolve any violations of model assumptions)
   3. Example 1: We checked the following four assumptions for MLR. First, we checked BLANK using BLANK (see Figure BLANK). Because BLANK, we believed BLANK is plausible, even though there is slight violations of BLANK. Second, we checked BLANK.
4. Discuss any limitations of your statistical analysis concerning what makes a review positive
   1. Example 1: We note that our model that we used to conduct tests may suffer from BLANK and BLANK, as evidenced by BLANK and BLANK

**Part 2: Recommendations for Businesses**

1. Based on your analytics you done above, concisely and clearly state precise recommendations for businesses. Again, this may overlap with the EDA section or Part I and if so I would try to avoid any redundancies.
   1. See examples from previous lectures for what constitutes concrete, precise recommendations for business owners.
   2. Make sure to discuss why you arrived at these suggestions with plots/tables/relevant stats.
      1. One suggestion here to better organize the business plan/recommendation would be for each suggestion/recommendation, have a paragraph or two explaining how you arrived at the conclusion.
2. Discuss any limitations of your suggestions/recommendations (or sometimes called “threats” to your analysis)
   1. Example 1: Our recommendations rely on some assumptions we have made about the data and the analysis. They are BLANK, BLANK, and BLANK. In particular, our model does BLANK and BLANK, brings some uncertainty in our results/interpretation in BLANK.
   2. Example 2: Our recommendations are likely robust even if the market conditions change because of BLANK and BLANK. However, if BLANK and BLANK were to have happened, then some of our results/suggestions, specifically concerning recommendations on BLANK may change to BLANK.
3. While this section looks small in this outline, it should be take up at least more than one page of the analysis.

**Conclusion/Discussion**: (summarize what you wrote above; final thoughts/discussions)

**Contributions and References**:

1. Example 1: JD reviewed/wrote/edited BLANK for report and BLANK for presentation. HK did BLANK, including BLANK. Overall, we met blank times, spent BLANK hours, etc.
2. Example 2: JD: Figures BLANK, model strength/weakness, final editing, presentation slides BLANK. HK: figures BLANK, data cleaning