Title

**Introduction**

(copy+paste)

**Exploratory Analysis**

* Plots , summary
* Discuss a bit outliers
* Data preprocessing: missing values, split the “name” into “brand”, based on the median of brand, create “brand\_level”

NOTE: plots here are messy, might place them in panel by ggplot2

**Statistical Analysis**

Guide:

* Multicollinearity: Based on corrplot in EDA, we check collinearity
* Inclusion of interaction of continuous covariates by comparison
* Observe diagnostic diagram, try log transformation
* Merge levels based on log model, and get log\_merge model
* Consider to add/remove interactions by procedure
* Compare two models
* Get the final model

NOTE: output the summary of model by some “packages”

**Discussion**

Ideas:

* Based on the final model, answer research question 1,2,3
* Summarize/interpret the new found model
* Extra assumption of normal distribution of error term, since Q-Q plot is NOT perfectly normal
* limitation: the outliers/large value observations are hard to handle, though we use log transformation to reduce the effects.
* etc.