1. Ace

Their project goal is stated very clearly. It is nice to use two layers of web scraping to collect data that answers different questions. Use different machine learning models including regression models and random forest.

The drawback was that they didn't talk about how they choose features. Also, the word didn't fit in the story line very well. They almost only use web scraping to collect data, could consider other sources of data as well. In addition, they used regression models most of the time, could utilize more other machine learning models.

2.Advertisement wanted

Using the daily topic to start the presentation is very engaging. The analysis is comprehensive, as they utilize several different machine learning algorithms such as RNN, text mining and so on. The sentimental analysis is clear. The recommendation system based on sentimental scores and likes/dislikes is a very impressive final product.

Besides the recommendation system, the final results and summary is not very clear. There could have been a clearer explanation of model results as well as the insights they gained from the models in the end.

3. Avengers

This team conducted very impressive feature engineering process. Explain why they use certain models (i.e. MSE) very well. Great graphic visualization.

Yet their presentation puts a lot of effort into technical aspects, and could have more interpretations of model results and relevant insights, which would be a more attractive presentation.

4. Damafia

The data visualization and data exploration part are good. They put a lot of effort in understanding and cleaning the dataset. The way they classify the severity of the accident is good. The goal is very clear and the final product have real-life application.

We noticed that the accuracy of the model is not very high. They could try different machine learning methods and find the most accurate one. Further feature engineering may help, too.

5. Data massaging

The topic is very unique and interesting. Clear visualization using pie charts and other tools to make sense of the data. Considered quite a few important features and showed very clear thought process.

Analysis is not enough. Didn't use enough machine learning models. May try to use a larger dataset with more features for machine learning. May also try to use different more sophisticated machine learning methods

6. Data Samurai

Topic chosen about president selection is a very interesting data mining topic. The step by step breakdown of the topic is great. Very good introduction to the topic. Stated clearly about their project goals. Use various models, text mining and visualization to support their conclusion. It makes it very clear to show the table of model performance comparison after modeling.

They use model to predict last election's result and didn't actually predict future potential result, which is different from their goal. Could add a part and use the best model to predict this year's result to give the audience a sense of what would the result look like for this year.

7. DDOG

They explained very clearly about how they clean the data (especially the location). In sentiment analysis, the way they simplify the method and save computational time is great. They also talked about the limitations of their model, such as failure to validate data, collect image data from twitter.

They could add more machine learning models in the future, and focus more on key insights they obtained during presentation.

8. Fantasy three

The goal is defined very clearly, which includes finding the relationship between budgeting and revenue for each movie. They have chosen suitable models to analyze that, using classification and regression models with relatively good predictive power.

The storyline of the slides could be more complete. Could enlarge the dataset and add more analytic models. Could find a clearer storyline and explore some real-world application of the final product.

9. Hello World

They have a very clear goal of their project: find out what kind of host is more popular. The data contains various information. They also use different visualization tools to show the difference between superhosts and other host.

The text mining part is not quite related to the theme. The price prediction models are all linear with not very good results. They could do a more granular classification in the future, to delve into what makes a host popular, instead of only having the two classifications: superhost and non-superhost. They could reduce the amount of codes on the slides and draw more business insights next time, such as suggestions for airbnb hosts. Also they could try more deep learning model and unsupervised learning model like clustering.

10. Pandas

Clear theme to analyze the what impacted the airbnb prices to inform customers choices. Besides airbnb dataset, external datasets such as transportation data are also included, which makes the analysis more complete. It is also very nice to use different machine learning models and compare the performances.

Manage the presentation time better next time. The data preprocessing could have done more thoughtfully, to select the crucial features into models only to improve model performances.

11. Suicide Squad

The topic is new and creative. It is clear and nice to illustrate and trade off the true negative rate, recall , accuracy rate and other performance indexes of their own models.

Could have managed the time in presentation better. Also it would make the presentation more interesting to talk more about impact and insights of their analysis, rather than talking too much about technical aspects.

12. Tofu Chilli

It is a very intriguing topic to uncover the NBA blackbox using data mining. They take into consideration many relevant aspects and build a fairly comprehensive datasets which include age, weight, public comment, amount of winning and losing games, and even us economics. It is also thoughtful to reflect on the limitations of their modeling.

Breaking the stereotype in the conclusion part which is supported by real life examples is good. Yet it would be even better to utilize the modeling results to draw on some interesting findings of their analysis. Also, they could try more ways to yield a better performance of the model, such as trying different models and parameters.

13. Uncle-Luoyang Everyday

They did a good job in data preprocessing and quantifying key variables, by transforming dirty data into sentiment score, popularity score and dummy variables of genres, as preparation for modeling. And it is convincing to compare model performances and choose Catboost out of the four models they have tried.

Yet, there is still space for model performance improvement in the future. They may perform feature engineering to further select import features to put into models. In the conclusion part, it would be more insightful to talk about specific suggestions for movie producers according to their analysis.

14. Untitled

They gave a clear explanation of how they processed the data and conducted their analysis. Specific examples of rules were demonstrated for better understanding. The final product of the recommendation system is very complete. Good visualization.

Could improve the model since the scores are not very high. Spend more time on feature engineering next time or try more machine learning models to achieve a more accurate prediction. Can conduct further analysis on recommendation based on the result.

15. Unreachable

They explained clearly their goal which was to use yelp reviews to find business insights and predict ratings. Nice to make use of web scraping. Great data exploration and feature selection using different visualization tools. They provided a clear timeline to conclude how they conducted this project. It is nice to talk about why they used specific model, such as NRC text mining to reach greater efficiency and include a larger variety of features, instead of only the positive and negative.

They could include more datasets of the restaurant conditions to yield more insights. May add more models in the future such as different regression models and clustering to have a better sense of what differentiates a restaurant among all features.