# Structure

**Abstract**: 1-2 paragraphs, executive summary

**Introduction:** what is the problem? why this problem? potential challenges, key ideas

**Related work:** what has been done in the literature? limitations? How is your work different?

**Methodology** (may break into multiple sections)

✦ Datasets, tools, main tasks, analytical thinking

✦ data understanding, preprocessing, warehouse, modeling

**Evaluation**: metrics, evaluation setup, baseline methods, results (figures, tables w/ proper labels), interpretation

**Discussion**: lessons learned, what worked well, what didn’t, directions for future work

**Conclusion**: summary, reiterate key tasks & findings

**References**: check format for proper citations

**Appendix**: Honor Code Pledge, individual contribution

# Abstract (Peter)

* To be written after all other sections are completed

# Introduction

* Probably copy from previous check-in report

# Related work (Peter, Vijay, Andrew, Harrison)

* [MetaQ] Are there any research which was not included previously in this section that you want to use to compare to the performance of something we built?
* What is the research done?
* Who did the research?
* When did the research get published?

# Methodology (Peter, Vijay, Andrew, Harrison)

## Datasets, tools, main tasks, analytical thinking (Peter, Vijay, Andrew, Harrison)

For each dataset (Vijay)

* What data are we collecting?
* How much data are we collecting?
* Where are we getting the data?
* Add citation for dataset sources

For each tool used (Peter, Vijay, Andrew, Harrison)

* What are we using it for?
* Are there alternatives?
* If there are alternatives, why did we pick the tool we are using?

For each main task (Peter, Andrew, Harrison)

* What is the main task?
* What was our initial approach to solving the main task?
* What difficulties did we find in solving the main task? How did these difficulties change our approach to solving the problem?
* What are the technical details of the method that we finally used?

Analytical thinking (Peter, Andrew, Harrison)

* What kinds of patterns did we initially expect to find?
* Why did we collect the data for specific stocks?
* How can we interpret the solution to each main task? (i.e. what is the practical implication of the solution)
* Are there any caveats or assumptions we need to make clear when giving the interpretations?

## Data understanding, preprocessing, warehouse, modeling (Peter, Vijay, Andrew, Harrison)

Data understanding (Peter, Andrew, Harrison)

* What exploratory analysis did we do on the data (what tables/graphs did we make that don’t directly answer the main questions, but still give interesting information)?
* Why did we do the exploratory analysis that was listed?

Preprocessing (Peter, Andrew, Harrison)

* Did we need to clean any data (replace missing values, substitute values, etc.)?
* Why did we make the choices we made when cleaning data?
* Did we need to perform any normalization or other transformation of the data?

Warehouse (Vijay)

* What does the raw data look like?
* How is the dataset stored?

Modeling (Peter, Harrison)

* What models did we make to answer the main questions?
* What inputs do the models take?
* What outputs do the models give?
* Why should the models give reasonable answers to the questions we are interested in?
* Did we make any assumptions about the form of the data which relate to the exact model we chose to use?
* What parameters of the model could affect the output of the model?

# Evaluation (Peter, Andrew, Harrison)

Metrics

* What are we measuring about the model – how do we know our model is good?
* What statistics can we state about the model quality?

Evaluation setup

* How do we train the models?

Baseline methods

* What prior work is comparable to our models?
* What simpler models are comparable to our models?
* What is the difference between our model, the prior work models, and simpler models?

Results (figures, tables w/ proper labels)

* What outputs did we get from the models?
* What input did we give to the model?

Interpretation

* What can we learn generally from the results we got from the model?
* Are there any specific, interesting, or surprising findings from our results?

# Discussions (Peter, Vijay, Andrew, Harrison)

Lessons learned

* Are there any techniques that you learned about through the project that you didn’t know about?
* Are there any significant data mining principles that you got to experience in the project firsthand?

What worked well/what didn’t

* Is there any technique or part of the project which worked particularly nicely in the project?
* Why do you think that it worked well? How could you use this knowledge in the future?
* Is there any technique or part of the project which worked particularly badly?
* Why do you think that it worked poorly? How could you avoid or improve upon the deficiencies?

Directions for future work

* What would you want to do if you had another month on the project?
* What ideas do you have for exploring the data which are not directly related to the techniques we used in this project?

# Conclusions (Peter)

To be completed after all other sections are written

# Appendix (Peter, Vijay, Andrew, Harrison)

We include the honor code pledge in the appendix to certify that we have neither violated nor concealed any violations of the University of Colorado – Boulder honor code.

``On my honor, as a University of Colorado at Boulder student, I have neither given nor received unauthorized assistance on this work. ‘’

Work done by each team member:

Peter: data understanding, clustering, sentiment analysis.

Vijay: Dataset collection, data preprocessing, database design.

Andrew: Evaluation, clustering analysis, model interpretation.

Harrison: Methodology, decision tree implementation, discussion.