# Diversity of Maintenance Logs and Delay

Bryan Anderson

13 May 2018

# Summary

- Introduction
- Motivation
  - Federal Aviation Administration (FAA)
  - Natural Language Processing (NLP)
- Data
  - Remote Monitoring Logging System (RMLS)
  - Bureau of Transportation Statistics
- Methodology
- Results
- Conclusion

# Introduction

- Bryan Anderson

- Consultant and Software Developer with Cobec Consulting

- B.A. in Economics and Mathematics from Augsburg College

- M.S. in Industrial & Systems Engineering from the College of Science & Engineering University of Minnesota – Twin Cities

- 7 years of Industry Experience
  - Industrial Engineering – Supervalu Inc

cobec
CONSULTING

# Federal Aviation Administration

- FAA is responsible for the safety and growth of civil aviation.

- There is an annual report to plan investment options to support the goal above called the Capital Investment Plan (CIP).

- The CIP requires several analyses to be conducted. Each of these analyses has varying level of detail depending on the maturity of the program. The maturity is defined by the FAA's Acquisition Management System (AMS) process.

  - Cost estimation is essential in the process.
  - Data is difficult to procure.

cobec
CONSULTING

# Natural Language Processing

- There is abundant of data for analyses, but the problem is it's not in a useable format yet.

- New technology and techniques are becoming more robust that enable the use of the previously un-useable datasets.

- One of these new techniques and technology is Natural Language Processing (NLP).

- NLP is a field of science focused with translating unstructured text into a structured format for analysis.

cobec
CONSULTING

# Remote Monitoring and Logging System

- RMLS is the FAA's maintenance logging system. Field technicians are alerted and tasked maintenance activity. After performing the work, a description of what was performed is logged.
    - Example case 1
    - Example case 2

cobec CONSULTING

# Bureau of Transportation Statistics

- BTS has rich datasets on all modes of transportation and topics.

- Dataset used in this report was the "Airline On Time Performance". Monthly-subset of domestic flights in the US. Information includes departure/arrival times, delay, and other flight metrics.

- The label used from this dataset was the arrival delay.

cobec
CONSULTING

# Feature Extraction

- Bag-of-words model approach.
  - Each word is feature in a the model.
- RMLS log data was aggregated by the airport-day. E.g., all logs from a given airport on a particular day are concatenated for one observation.
- Airport-day observations were then tokenized into sentences, and then into words.
- Remove stop-words and common punctuation characters
  - Isn't, won't, very, just
  - , . ] [ ) (

Proprietary Information

cobec
CONSULTING

# Feature Extraction

- Evaluated Term Frequency – Inverted Document Frequency (TF – IDF) on each airport's day to measure the importance of words in the text.

- The mean of the TF-IDF vector was then used in order to compare facilities on the diversity of their log messages. The more 'unique' log messages, the higher the average is.
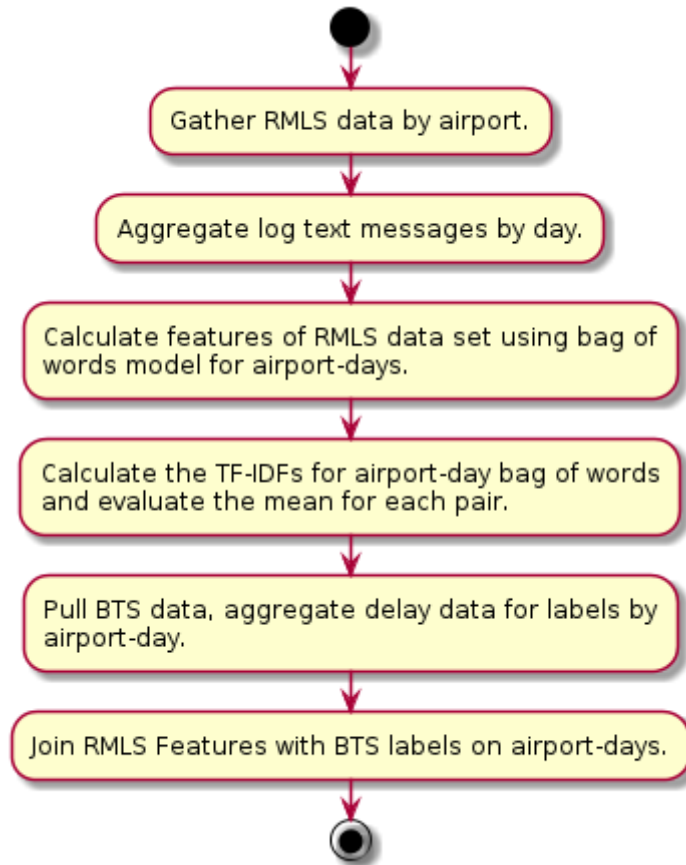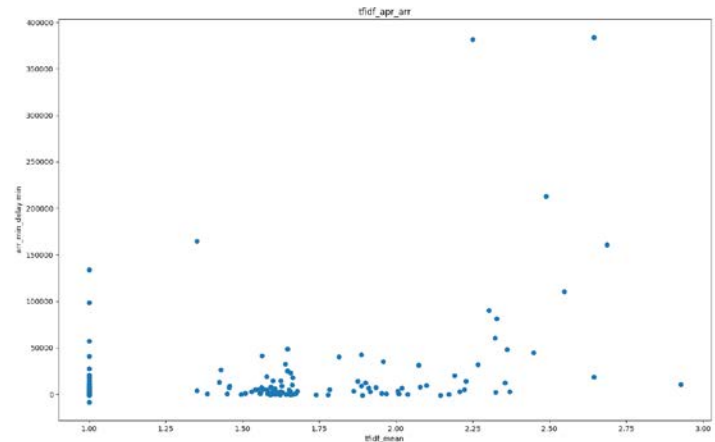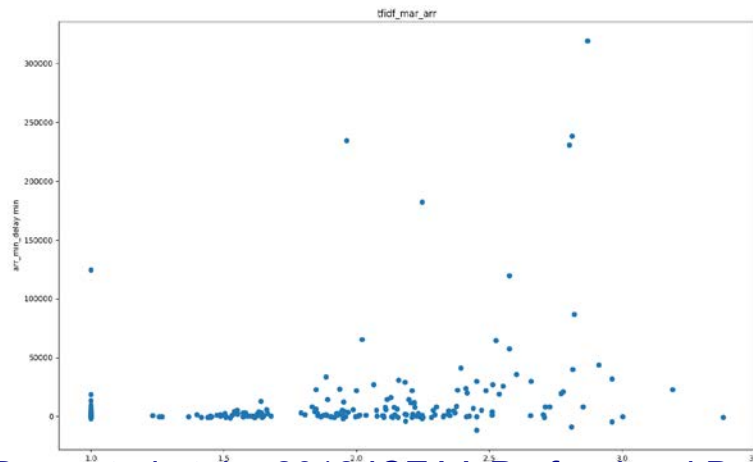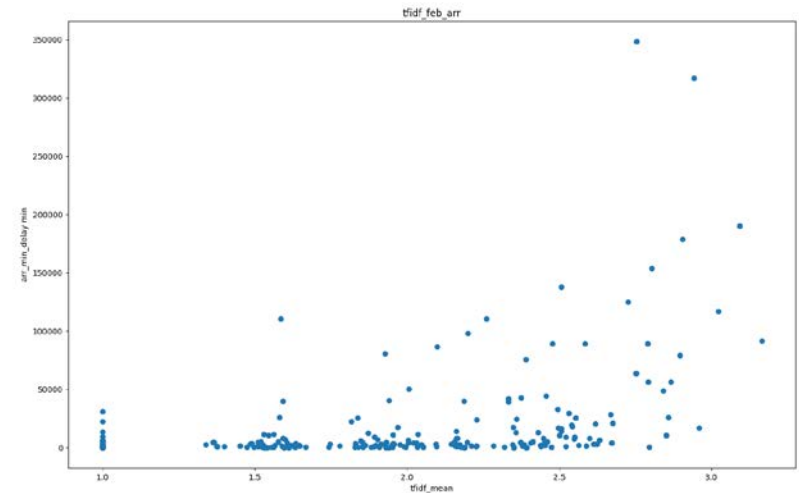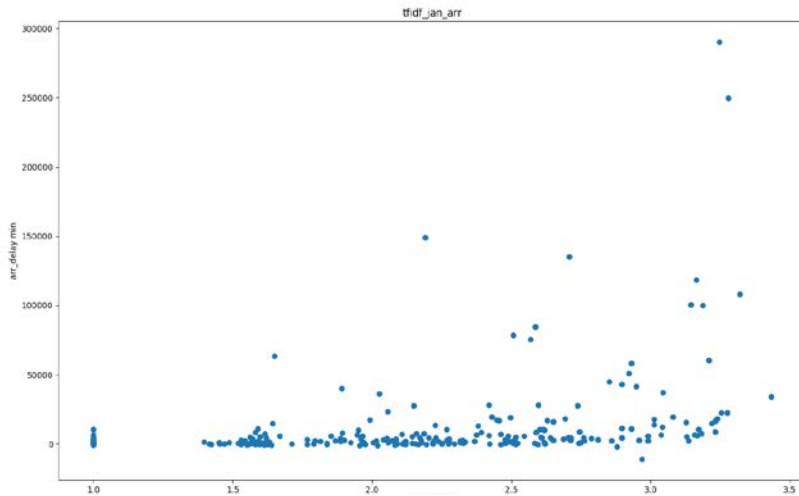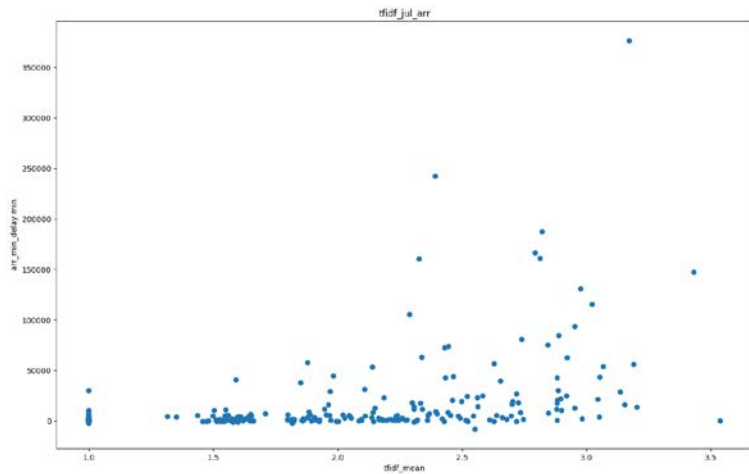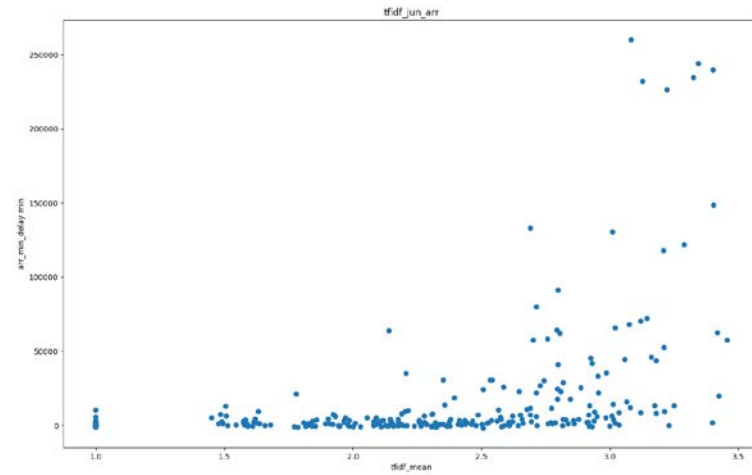
cobec
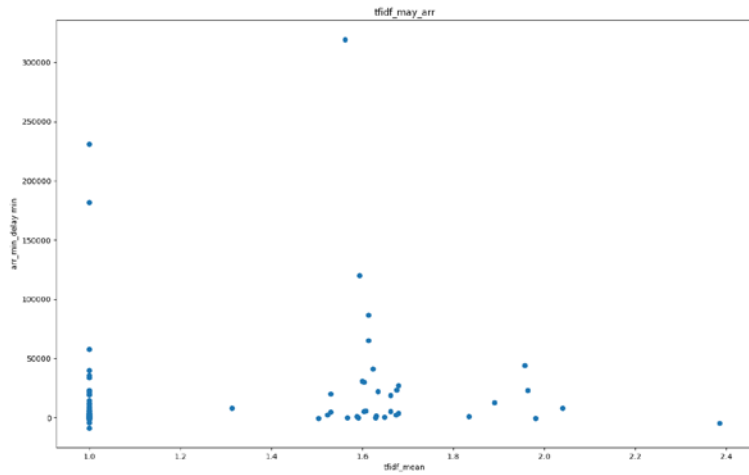CONSULTING

# Methodology

10
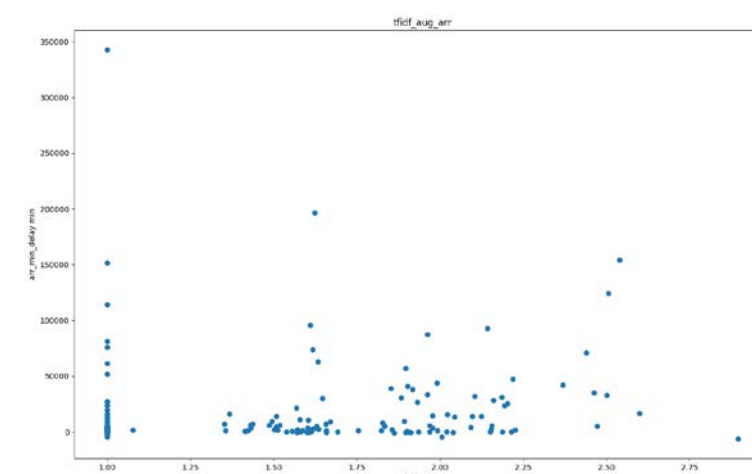
# Results

# Results

# Results
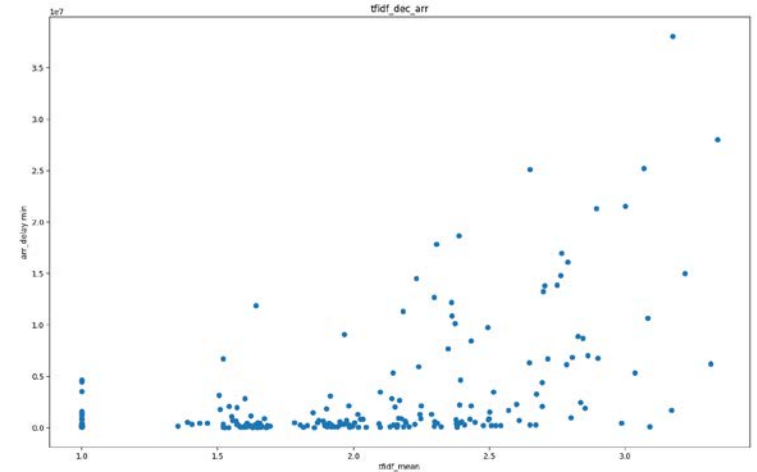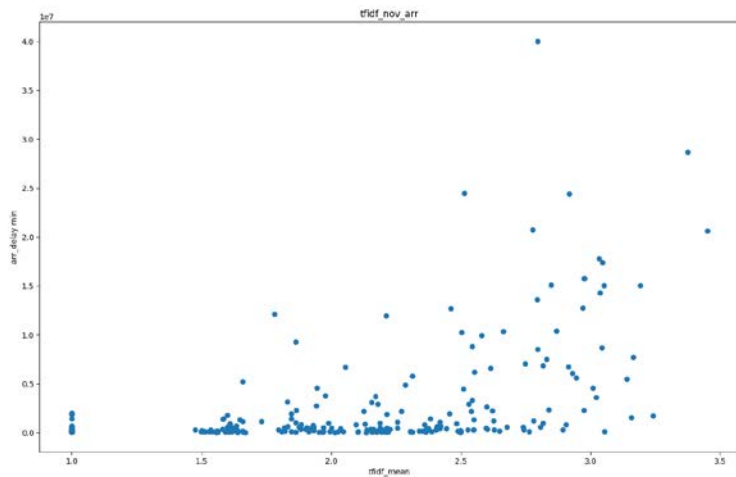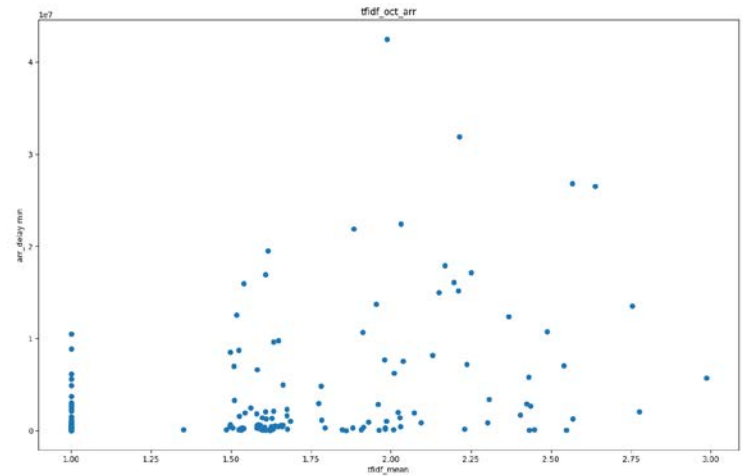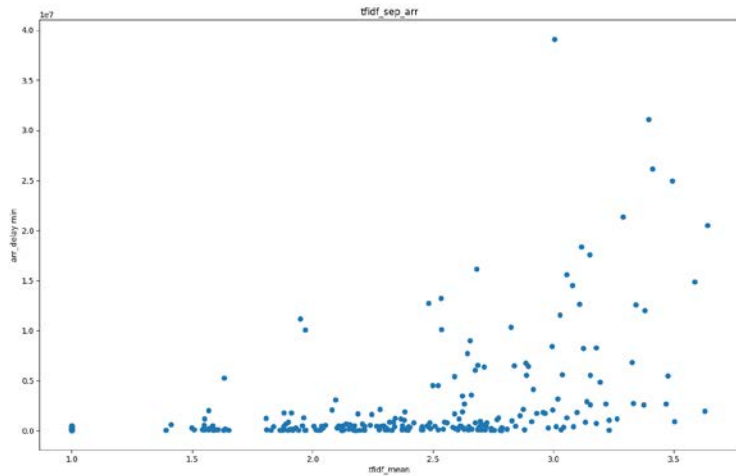
# Conclusion

- The general trends indicate higher average TF-IDFs correlate with higher delay.

- Understanding why higher average TF-IDF vectors could be useful when developing LCCEs.

- Next steps are too look at different techniques to identify high frequency key words.

cobec
CONSULTING

# Questions?

cobec
CONSULTING