West Texas A&M

BI Roadmap

Airline Service Evaluation

Date: July 2023

Author(s): Jason Hardy

Sponsor(s): CIDM-5310

TABLE OF CONTENTS

BI ROADMAP - SUMMARY

Table 1: BI Roadmap Highlights

Project Release	N	N + 1	N + 2	N + 3
Timeframe	7/9			
Key Deliverables	7/23			
Business Applications	7/16			
Data Subjects	7/14			
Source Systems	7/11			

BI RELEASE: FINAL PROJECT

Timeframe

1. Identified data sources – 7/1 - 7/9
2. POC of data pulls – 7/7 – 7/9
3. Setup virtual environment – 7/9 - 7/11
4. Setup host – 7/9/ - 7/11
5. Setup data infrastructure – 7/11 – 7/13
6. Data transformations – 7/9 – 7/13
7. Initial data loads – 7/13
8. Data story and Essay – 7/23

Key Deliverables

- Business Deliverables

  The ability to demonstrate various levels of information about airlines.

  Specifically I wanted to provide (as a first phase) a coorilation between airline profits and their customer service level in terms of lost baggage and oversold flights.

Future deliverables such as fares for various cities are a part of the design, and data was imported, but hasn't been actioned upon yet.

- Data Deliverables

I am looking to provide details on:

1. Airline profitability and customer service levels
2. Customer mitigation of poor experiences (e.g., delays, cancellations, bags mishandled, etc) by evaluating the effect on revenue over time.

- Sourcing Deliverables

Utilize data sources from the Bureau of Transportation. Specifically, the following sources hold potential information:

Commercial Aviation - Involuntary Denied Boarding

Commercial Aviation Mishandled Baggage and Mishandled Wheelchairs

Consumer Airfare Report: Detailed Fare Information For Highest and Lowest Fare Markets Under 750 Miles

Airline Quarterly Financial Review - Majors

- Technology Deliverables

Provide a series of dashboards that demonstrate how the data correlates.

Utilize Pandas along with Juypter as the main mechanism for serving the data and have a cloud-based service that can handle the serving of the information. I am specifically looking to provide charts that will help explain any correlations (or lack thereof). Specifically, I think the matplot will aid in this. I also believe the series capabilities of

Provide a database that will contain the information pulled from the various sources

Have automated jobs that will query the APIs as needed to obtain up-to-date information

Business Applications

- Pandas

- Jupyter
- Sodapy (library for communicating with Socrata, BTS.govs hosting service)
- MySQL

Source Systems

1. Bureau of Transportation APIs via Socrata.

Technology Required

- New
    - Virtual Server/Cloud service to host
    - New databases to store the data from the APIs