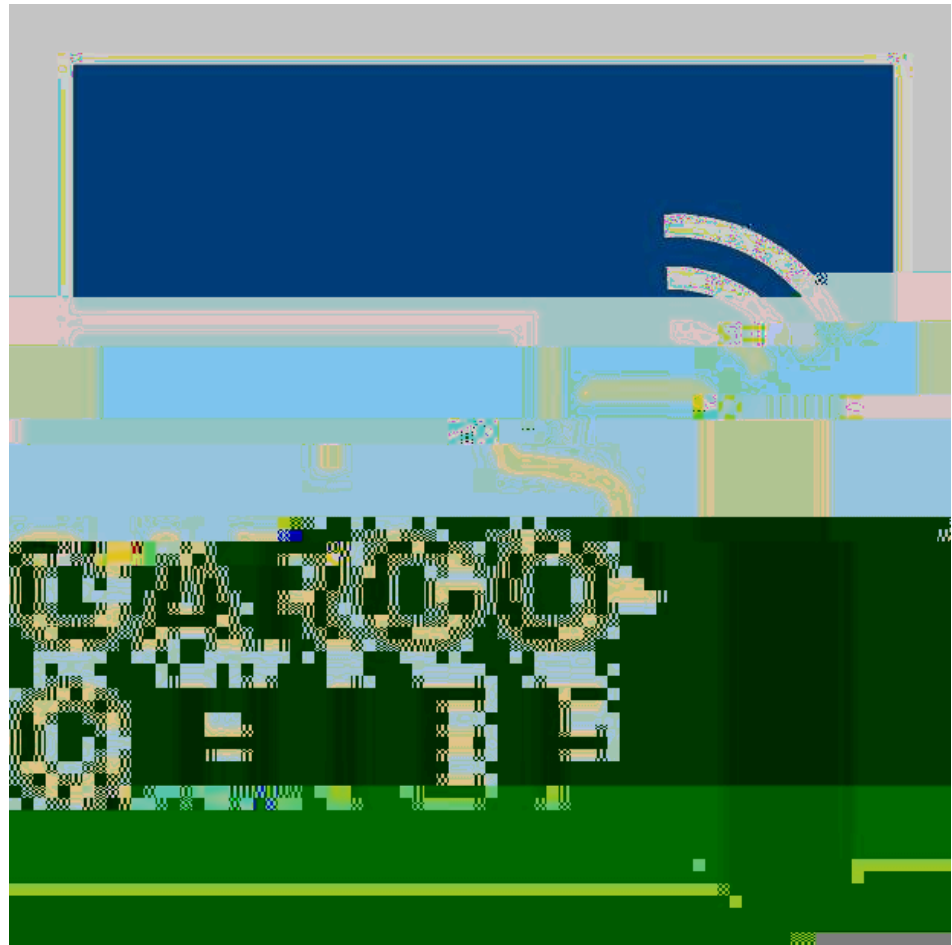! " #$%&' ( )%

#" * +, %! " - *)%

! 2%)1#+*1-$)3%

Company: Cargo Chief, Inc.

Employees: 50-100

Position: Data Science Intern

Supervising Statistician:
Tracy Holsclaw, Ph.D.

– ⁄ 45+6#%! " ##+*%

SHIPPER: someone with stuff to send from point A to point B

CARRIER: someone capable of transporting stuff between points A and B

BROKER: intermediary that connects shippers and

TRAILER:

3*. 40+! %6. 1#+7#%

Problem: For a given lane provide Cargo Chief a cost (paid to carrier) and a price (charged to shipper).

Inputs: Origin, Destination, and Trailer

Outputs: Cost and Price

Data Sources: Proprietary and third-party.

Model Overview

1. **Redefine lane as the unique combination of Origin, Destination, and Trailer.**

2. **For each lane provide a set of coefficients to predict cost.**

3. **Add a margin to the estimated cost to determine estimated price.**

4. **Updated coefficients every week.**

Autoregressive process (AR): account for lane-specific history

Multiple Linear Regression (MLR) : reinforce estimates with general patterns

Spatial Spline: construct an integer-valued predictor variable for MLR

Model Averaging: combine AR and MLR when necessary

Numerical optimization: determine ideal weights for AR and MLR estimates.

0" 19/" 9+-%

SQL

Query for the data

R (R-Studio)

Built an R package specifically for our algorithm

Data Cleaning

Model Fitting

Writing output files for implementation

General data analysis

Automated reports with Sweave

PHP

Implement algorithm on live server

60. -)19%#$. /9$#-%

Data science in industry

Information vs. Data

Talk to anyone and everyone

Start-up environment

Machine Learning vs. Statistics

Q & A