Team DAY Project

Dustin Bracy, Adam Ruthford, Yang Zhang

1/23/2020

## Introduction

We would like to answer the following questions about our data:

1. Can we accurately predict the runtime of our reports?
2. What is the expected average delivery time of a one off report?
3. What are the best & worst times of day, week, or month to run a report?
4. Is there a difference in performance between report servers?
5. What are the biggest impacts on performance of reports?

# Data Description

We have collected report data spanning January 2019 through January 2020. The available parameter list is very large, having over 500 parameters used. In order to narrow the scope of our project, we focused on completed reports only, and determined the parameters show below would be most useful in predicting report run time. A description of utilized fields is avaiable below:

### Report Metrics

* Priority - controlled report parameter
* QueuedDateTime - when report entered queue
* RenderStartDateTime - when server began work on queued report(s) in queueID batch
* ReportStartDateTime - when server began work on individual report
* ReportEndDateTime - when server completed work on individual report
* RenderEndDateTime - when server completed work on queued report(s) in queueID batch
* ReportBytes - size in bytes of completed report
* RptFormat - format of the report delivered
* SchedFreq - Scheduled Report Frequency
* Server - Server on which the report was built

### Engineered Features (all time is in seconds)

* AgentCount - number of agents selected
* DayOfMonth - Day of the month report ran
* DayOfWeek - Day of the week report ran
* GroupCount - number of clients selected
* HourOfDay - Hour of the day report ran
* LagTime - Time report waited for batch to finish (RenderEndDateTime less ReportEndDateTime)
* QueueTime - Time report entered queue (RenderStartDateTime less QueuedDateTime)
* ReportBuildTime - Time spent to build the report (ReportEndDateTime less ReportStartDateTime)
* ReportDeliveryTime - Time user spent waiting for report (RenderEndDateTime less QueuedDateTime)

# Exploratory Data Analysis

# Appendix

## Additional Report Field Descriptions

* Agents - Individual AgentIDs
* CurrencyCode - Collection of different lines of business
* GroupCustom - Collection of different client types
* GroupFI - Fully Insured Client
* GroupSF - Self Funded Client
* ReportID - Uniquely identifies a report
* QueueID - Uniquely identifies a queue of one or more reports
* FormID - Uniquely identifies collection of report parameters

## Report Category Crosswalk:

**Eligibility** “Eligibility”,“Accumulator”,“EventHistory”,“COBRA”,“RedCard”,“ComPsych”

**Miscellaneous** “Miscellaneous”

**Reinsurance** “Reinsurance”,“Aggregate Report”

**Professional Liability** “PL”

**Actuarial** “Actuarial”,“ISBReports”,“Renewal”, *except when rptCategoryDesc = ‘Renewal’ and reportGroup = ‘CommissionsTFB’ use ‘Commissions’*

**Finance Commission** “Commissions”,“CRMOnlineFinance”,“CommCnt”,“CommCompare”

**Claims** “ClaimsAudit”,“CMSRepting”,“Repository”,“GilsbarPPO”,“Claims”

**Medical Management** “Wellness Repository”,“MedCom”,“TVC”,“MedInsight”

**Metrics** “iTrac”,“TMS”,“Customer Service”,“ODR Admin”,“PortalStats”,“AutoAdjudication”,“Genelco”,“Supervisor”

**Extract** “DataExtracts”,“Large Claims Reports”

**Finance Billing** “Billing”, “Coverage”,“Refunds”,“Credit Card Process”,“Deposit”,“CheckRecon”,“Check Register”,“FinancePremium”, “Premium”

## ETL of original data: