1. Overview:

The Data Science team is working to create a predictive model for Claimant Motivation to RTW to augment Medical condition, Functional Capacity and Vocational Assessment. While the model predicts motivation scores of claimants, rules to operationalize motivation score along with other factors will help LTD and Clinical drive appropriate action plans

Identify information to capture to create a pool of relevant Motivational data for use to enhance the Model over time and subsequently leverage for Motivational insights.

2. Scope

The scope of the FRD is to document the ETL process for LTD Motivational Analytics. The scope of the document only limits to the data preparation process and post processing of model results for LTD Motivational Analytics.

3. Historical load process

Not applicable

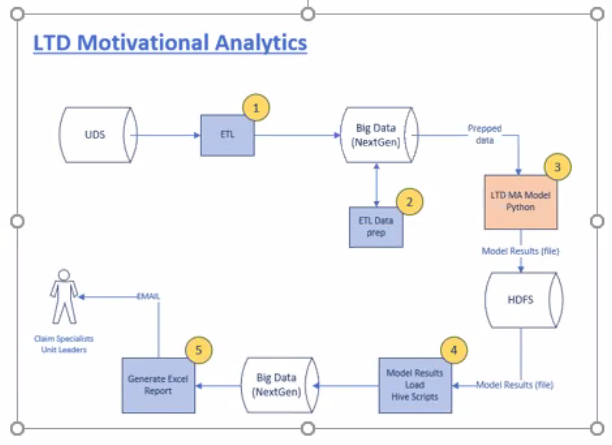
4. Technical Assumptions, Dependencies, Constraints and Risks

Technical Assumptions

Dependencies • UDS data refresh before the LTD Motivational Analytics process begins.

Risks • If there is any change in structure of the results file, it will impact the ETL process

5. Process Flow



6. Functional Requirements

6.1. Data Extract/Load

The below tables are expected to be refreshed from UDS, DRRM\_DM into NextGen PROD before the LTD Motivational process begins for the current day processing. Following is the list of tables that are required for LTD Motivational analytics scoring and report.

UDS (covered by the UDS extract ETL job):

1. TUDCLINF

2. TUDCLMBR

3. TUDEMPLY

4. TUDTRTMT

5. TUDCCLM\_PRGRS

6. TUDDPDNT

7. TUDEMPLR

8. TUDPLPRV

9. TUDPLPV2 (to be included in the UDS extract process)

10. TUDCAL10 (to be included in the UDS extract process)

11. TUDBN (to be included in the UDS extract process)

Disability DataMart (DRRM\_DM production):

1. T\_ICD9\_LKUP (already available in Analytics db in NextGen)

The below tables are one time loads to be done from csv files. These CSV files are available in the NAS path below.

\\Ns-sisci01\ghcmo\Projects\Motivational\

6.2. Data Prep (Model Input)

The Data prep involves creating a scoring set table. The scoring set table which will serve as the input dataset to the model.

• Only the claims which satisfy the below conditions must be selected for scoring.

 if Claim type/coverage = Long-term disability (LTD) then (if CLM\_NUM\_CD is > '600000000000' then LTD) and

 if Claim is on or after the 45th day from the received date of the claim and

 if Claim has at least one record in the treatment table “tudtrtmt” and

 if total count of valid ICD values present in the treatment table is at least 1

 if the claim is not already scored (If the claim is not present in the T\_HIST\_LTD\_MOTV\_ANLT\_SCOR\_SET table)

• Target scoring set should contain 1 record per claim number.

• LTD claims that have already been scored by the Model before must be excluded.

• Historical data of the scoring set must be maintained in hive table T\_HIST\_LTD\_MOTV\_ANLT\_SCOR\_SET.

• Scoring set table: T\_LTD\_MOTV\_ANLT\_SCOR\_SET – Truncate & Load

Please find the table structure and mapping document attached in section 10.1 LTD\_MA\_Data Description document.xlsx

6.3. Model Results Load

The model results will be available in pipe-delimited text file from the scoring process. Refer to the data description document (Model Results tab) in Appendix section for the list of fields and field order. (To be confirmed) All the highlighted portion below are to be decided and confirmed by DS team

o Results file location: Data/analytics/ghcmo/model\_results/ltd\_ma\_results\_set\_YYYY\_MM\_DD\_HH\_MM\_SS.txt (file name to be the same as the table name)

o Ignore header row in the results file.

o If case of multiple results files being present, all of them must be loaded in order of date/timestamp.

o After loading the results file to hive table, it must be moved to Data/analytics/ghcmo/model\_results/archive folder.

o Historical data of the results set must be maintained in hive table.

o Results table: T\_LTD\_ MOTV\_ANLT\_RSLT\_SET – Truncate & Load

o Results staging table : T\_STG\_LTD\_ MOTV\_ANLT\_RSLT\_SET (all fields must be string) – Truncate & Load

o Results history table: T\_HIST\_LTD\_ MOTV\_ANLT\_RSLT\_SET – Append only

6.4. Model Report

Generate a LTD Motivational Analytics report (text file) based on current day’s model results. Refer to the attached Report Mapping document for the list of report fields and field order.

Report Structure attached in section 10.1 LTD\_MA\_Data Description document.xlsx.

o Claim # is the key field. The report must contain only one record per claim #.

o Report should include all claims from the current day’s result set.

o Report File Name: LTD\_Motiv\_Analysis\_report.xls

o Report Distribution by email, to <To be provided> DL to be used only in PROD environment Dev/QA: AD team & Deepan –Use the same email template that we have currently. Sample email attached in appendix.

o Report Table - T\_LTD\_ MOTV\_ANLT\_RPT – Truncate & Reload

o Historical data of the report must be maintained in a separate hive table - T\_HIST\_LTD\_ MOTV\_ANLT\_RPT – Append

7. ETL Target DDLS

Number of Tables:

<<To be attached>>

8. ETL Schedule

Expected SLA for the report is 8am ET.<<TBD>>

9. ETL Environment Details

9.1. ETL LDAP Group

Environment LDAP Group

DEV APP\_11469\_ANALYTICS\_BSG\_DEV\_RD,

APP\_6942\_GHCMO\_DEV\_RD

QA app\_11469\_analytics\_etl\_qa\_rd

app\_11469\_analytics\_etl\_qa\_ch

APP\_6942\_GHCMO\_QA\_RD

PROD app\_11469\_analytics\_etl\_prod\_rd

app\_11469\_analytics\_etl\_prod\_ch

APP\_6942\_GHCMO\_PROD\_RD

9.2. Source & Target system details

• Source UDS tables to be loaded in big data, NextGen cluster, uds\_hist database. In QA, uds\_hist\_qa will be the database.

• All other tables (lookups, scoring and results) will be loaded in big data, NextGen cluster, Analytics database. In QA, analytics\_qa will be the database.