Sticky Notes

Thinking about two different tracks:

1) 45min-1hr each week during lunch hours. Go through a free Udacity course or low pressure Kaggle thing that is low overhead and requires no real outside work from Bob and I, but people can still learn from.

2) 2 hrs per week. Work on more serious and higher level concepts for a small group of hardcore enthusiasts. Any one is welcome, but this is going to be much more serious. Serious black belt ninja stuff; we training killers over here.

SCRATCH PAD:

1) Links to Udacity course, Titanic Kaggle tutorial.

2) Some introductory web site that says what data science actually is.

3) Links to some podcasts and web sites. E.g. Data Skeptic, O'Reilly, TWIML, DataFramed, Talk Python to Me, etc.

4) Some recommended reading (Think Stats, etc.).

5) Some good you tube introductory stuff.

-New opportunities offices in Waterbury (main), Meriden and Torrington.

-Brute force approach to "fixing" city and zip code data. Code included in 'Databridge DB -TempAddressFix' script. Scanned for incorrect spellings, some research on addresses for incorrect zip codes. Not perefect, but "good enough".

-------------------------------------------------------------------------------

--Research:

Additional tables for Predictive variables:

HouseholdServiceLog

PersonEngagement

PersonHealthInsurance/PersonHealthInsuranceHistory

PersonPlan

PErsonPlanService/PersonPlanServiceHistory

PersonPreAssessmentDimension

Bin income?

AgencySystemPersonHousehold and ref\_RelationshipType to get more houshold info

IsHomelessOrEmergencyShelter. Check distributions of DwellingType to see if we want flags around Homeless and Emergency Shelter.

Potential Targets:

FctPersonAssessment/PersonAssessment

FctPersonAssessmentBaseline

PersonPlanServiceOutcome

Census/external data? Matching zip codes to set indicators like Low income neighborhood, etc

--Modeling alterations:

Binning of categorical variables

Creating dummy variables

Create training set

--------------------------------------------------------------------------------

CMS Dashboard:

1-3 are Person-centric fields.

4 is more household cenered.

1,2 and 4 may be most ready/biggest bang for the buck.

Services have outcome indictors >>

Outcomes are associated with NPIs >>

NPIs roll up to an NPI Outcome Indicators >>

Outcome Indicators roll up to a Goal

-For now, NPIOutcomeCodes in ref\_NPIOutcome beginning with 1 and 6.4

ROMA's report looks at a person's score coming into a time period and their score leaving their time period, not an average.

-------------------

DATABRIDGE DASHBOARD:

--------------------------------------------------------------------------------

CONNECTIONS:

CMS Remote Connection:

Remote: 173.162.231.201

Username: NEWOPPORTUNITIE\\smc

Password: Smc201!

Database -Test: NewOpportunities\_CMS\_Test.

Database –Prod: NewOpportunities\_CMS

SMC Network Servers:

smc-sql-2014,1434\\phi

smc-sql-2014

smc-sql-2008

smc-sql-2005