**POINTS TO BE CONSIDERED WHILE DOING INSIGHTS**

1. These two are related rows

|  |  |  |
| --- | --- | --- |
| 8 | AMT\_INCOME\_TOTAL | Income of the client |
| 9 | AMT\_CREDIT | total amount of money that is borrowed by the client |

1. This row is applicable only for consumer (goods price)

Try doing a parallel study

|  |  |
| --- | --- |
| AMT\_GOODS\_PRICE | For consumer loans it is the price of the goods for which the loan is given |

1. Try to build a relationships

|  |  |  |
| --- | --- | --- |
| 8 | AMT\_INCOME\_TOTAL | Income of the client |
| 12 | NAME\_TYPE\_SUITE | Who accompanied client when applying for the previous application |

1. To be converted into years

|  |  |  |
| --- | --- | --- |
| 18 | DAYS\_BIRTH | Client's age in days at the time of application |

1. RISK BASED ON ANNUAL (FRESHERS MIGHT BE ON NOTICE PEORID)

|  |  |  |
| --- | --- | --- |
| 19 | DAYS\_EMPLOYED | How many days before the application the person started current employment |
| 10 | AMT\_ANNUITY | Amount paid / year |
| 8 | AMT\_INCOME\_TOTAL | Income of the client |

|  |  |  |
| --- | --- | --- |
| 9 | AMT\_CREDIT | total amount of money that is borrowed by the client |

1. HOW TO DEAL WITH THIS ? (QUESTION FOR MENTOR)

|  |  |  |
| --- | --- | --- |
| 24 | FLAG\_EMP\_PHONE | Did client provide work phone (1=YES, 0=NO) |
| 25 | FLAG\_WORK\_PHONE | Did client provide home phone (1=YES, 0=NO) |
| 26 | FLAG\_CONT\_MOBILE | Was mobile phone reachable (1=YES, 0=NO) |
| 27 | FLAG\_PHONE | Did client provide home phone (1=YES, 0=NO) |

1. RISK ANALYSIS

|  |  |  |
| --- | --- | --- |
| 41 | ORGANIZATION\_TYPE | Type of organization where client works |
| 8 | AMT\_INCOME\_TOTAL | Income of the client |
| 9 | AMT\_CREDIT | total amount of money that is borrowed by the client |
| 10 | AMT\_ANNUITY | Amount paid / year |
| 19 | DAYS\_EMPLOYED | How many days before the application the person started current employment |
| 13 | NAME\_INCOME\_TYPE | Clients income type (businessman, working, maternity leave,…) |

Consider only workers working for any organization for this analysis

1. CONSIDER ROWS FROM 45-95 and conduct a feature engineering basedon below row conditions only

|  |  |  |
| --- | --- | --- |
| 16 | NAME\_HOUSING\_TYPE | What is the housing situation of the client (renting, living with parents, ...) |

1. Does population affect on these factors

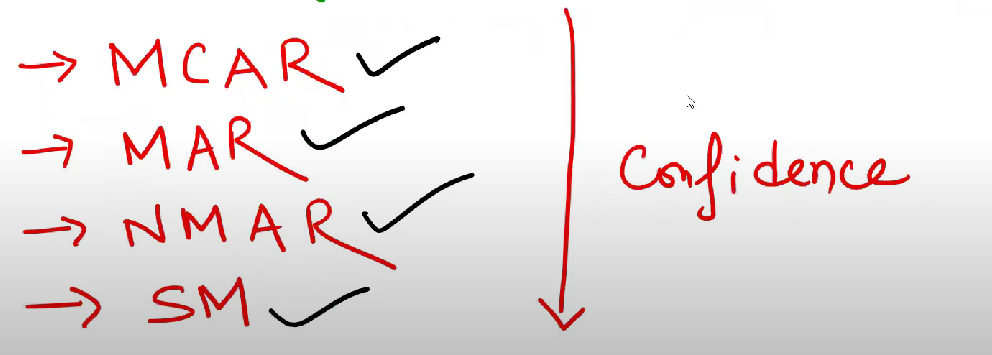
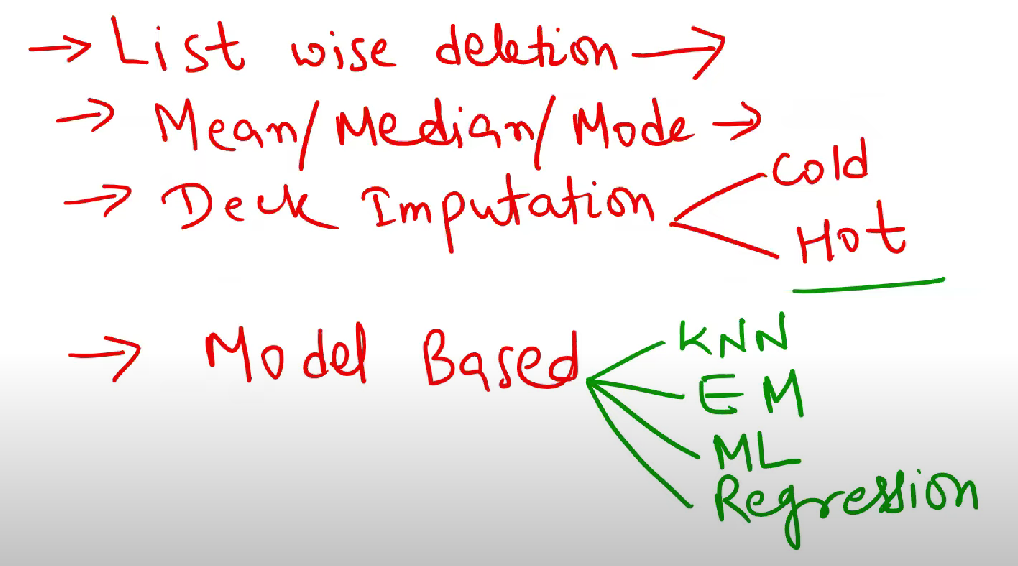
|  |  |  |
| --- | --- | --- |
| 17 | REGION\_POPULATION\_RELATIVE | Popoulation normalized value, higher value indicates more population in the area |
| 6 | FLAG\_OWN\_REALTY | Does the client has a property or not |
| 8 | AMT\_INCOME\_TOTAL | Income of the client |
| 16 | NAME\_HOUSING\_TYPE | What is the housing situation of the client (renting, living with parents, ...) |

1. Check these are duplicate columns or not

|  |  |  |
| --- | --- | --- |
| 35 | REG\_REGION\_NOT\_LIVE\_REGION | Flag if client's permanent address does not match contact address (1=different, 0=same, at region level) |
| 36 | REG\_REGION\_NOT\_WORK\_REGION | Flag if client's permanent address does not match work address (1=different, 0=same, at region level) |
| 37 | LIVE\_REGION\_NOT\_WORK\_REGION | Flag if client's contact address does not match work address (1=different, 0=same, at region level) |
| 38 | REG\_CITY\_NOT\_LIVE\_CITY | Flag if client's permanent address does not match contact address (1=different, 0=same, at city level) |
| 39 | REG\_CITY\_NOT\_WORK\_CITY | Flag if client's permanent address does not match work address (1=different, 0=same, at city level) |
| 40 | LIVE\_CITY\_NOT\_WORK\_CITY | Flag if client's contact address does not match work address (1=different, 0=same, at city level) |

Why there is a missing value check wrt to the feature

ADD info about what kind of missing – MNAR , MAR etc - <https://www.youtube.com/watch?v=-uC79UTOye8>

ASSESSING impact of imputation - **Pattern of Missing Data:**

**Check feature with highest missing value and find ithe highest correlated features to it and and find patterns**

Deleting rows or columnswithmissingdata is awidely usedapproach.However,asitwillbeillustrated, if thisnaive [Snapshotof theincompletedataset.] [Fullviewof thedataset, emptycells inblack.] Fig.1. Incompletedatasetderivedfromarealmanufacturingprocess. method isapplied,oftentimes thesamplesizemaybesignif icantly reduced, leaving little information for the algorithm to learn thepatterns.Thismethod isnot advisedunless the proportionofeliminatedrecords isverysmall (<5%)

a new greedy-like algorithm

reducing no of records might lead to bias error