Data Dictionary

For

**Source One Financial**

**Defaulted Loans Comparisons**

|  |  |
| --- | --- |
| **Date** | **Description** |
| 2019/6/24 | Initial Version  Data dictionary of all table columns except Vehicles. |
| 2019/7/7 | 1st Review.  Define ETL steps for data cleanup.  Review ER Diagram. |
| 2019/7/8 | Data Dictionary for Vehicle table.  Identify source of each table. |

Contents

[ER Diagram 2](#_Toc13379075)

[CreditDecision 3](#_Toc13379076)

[Customer 6](#_Toc13379077)

[Decisions 10](#_Toc13379078)

[IncomeData 13](#_Toc13379079)

[LossDetails 13](#_Toc13379080)

[Losses 14](#_Toc13379081)

[Vehicles 14](#_Toc13379082)

[Data on Losses - Quarterly LPC spreadsheets 16](#_Toc13379083)

[Data on Losses - Quarterly Additional recoveries 17](#_Toc13379084)

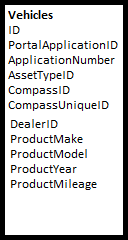
[Appendix A – VBA DispositionDate 18](#_Toc13379085)

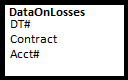
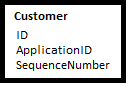
[Appendix B – Linking Defaulted Loans to Customer tables 19](#_Toc13379086)

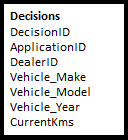
# ER Diagram

This section shows a condensed ER diagram, with each table only containing what appears to be key fields or any other fields required to join each table. This is to illustrate how each table relates to each other.

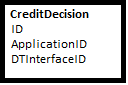
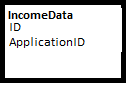
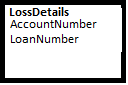
For our analysis we only require the following tables: Customer, Decisions, Vehicles and DataOnLosses.





The following three tables are not required for our analysis: CreditDecision, IncomeDate and LossDetails.

Each table includes only the fields that appear to be key values.



# CreditDecision

PK: ID

Rows: 574,254

Duplicate Rows: None

This table appears to be very similar to Decision table.

|  |  |  |
| --- | --- | --- |
| Keep? |  |  |
| \* | ApplicationID |  |
| \* | DTInterfaceID |  |
| \* | DecStatus | This value also appears in Decision Table. Is it the same?  AD  AP - Approved  BK - Booked  CA - Conditional Approval  CP - Credit Pending  DE - Declined  DI  DR - Document Received  EA - Expired Application  PP - Payment Pending  WD – Withdrawn  ? What is the code of Deficient Pending  See also CreditDecisionDescription |
|  | DecDateTime | Date and time |
| \* | Amount | Total Amount of loan |
| \* | InterestRate | Interest rate of loan  [blank], 0, 1, 2, 3, 4  And other values from 4.69 to 39.99 |
|  | Term | Duration of loan in months |
| \* | Amortization | Total amortization duration of loan in months |
| \* | TotalMthPay | Amount. Total monthly payment |
| \* | PaymentFrequency | How many payments will be made each year?  4 – once every 3 months (quarterly)  12 – once every month (monthly)  24 – twice a month (bi-monthly)  26 – once every 2 weeks (bi-weekly)  52 – once very week (weekly) |
|  | FirstPaymentDate | Date of 1st payment |
| \* | CashDown | Amount. Down payment on vehicle purchase. |
|  | LenderFee | Amount. |
|  | LicenseFee | Amount. |
|  | AdminFee | Amount. |
| \* | InstallPayment | Amount of each individual payment (not monthly amount, see PaymentFrequency) |
|  | OtherTaxable | Amount. |
|  | Reserve | Amount. |
|  | OtherNonTaxable | Amount. |
|  | DeferredInterest | ? all values blank or 0 |
| \* | AutomaticDecision | Flag: 0 or -1. |
|  | CashSalePrice | Amount of sale of vehicle |
|  | Installation\_Delivery\_Fee | Amount. |
|  | GPS\_Fee | Amount. |
|  | PPSA\_Fee | Amount. Personal Property Security Act ? Lien amount on vehicle? |
|  | CreatedBy | Internal user ID? |
|  | CanadianBlackBookDataID | ID to external table of Vehicles – Canadian Black book. |
|  | BureauDataID | ID to external table of Credit Bureau? |
|  | Tax | Provincial tax rate (GST?) |
|  | TradeInAllowance | Amount for trade-in |
| \* | TradeInDebt | Amount of debt remaining on trade-in |
|  | Rebate |  |
|  | ExtendedWarranty | Amount for extended warranty purchased. 0 if none. |
|  | OptionalLifeInsurance | Amount for optional life insurance. 0 if none |
|  | OptionalAHInsurance | Amount for optional for accidental health insurance. 0 if none. |
|  | OtherCost | Amount |
| \* | ApprovalReason1 | As per Credit Supervisor  Conditional Approval  Credit Pending  Fits SOFC Policy  Manager Exception  To Win from Competitor  Well Paid SOFC auto |
| \* | DeclineReason1 | Arrears in Last 6 Months  Auto Write-off in Last 24 Months  Budget Decline  Credit Pending  Decline condition 3  Decline condition 4  Decline condition 5  Declined at Income  Doesn't meet min. score  In BK/Consumer Proposal  Policy Break  Recent arrears in last 6 months  Shopping/Declined Customer  Suspect Fraud  Suspected fraud |
|  | InsuranceTax | Tax rate on insurance |
|  | ExtendedServiceContractTax | Tax rate on extended service |
| \* | RiskTier | See also PriceToRiskRate  S1  S2  S3  S4 |
|  | BaseReserve | Amount between 0 and 350 |
|  | AdditionalReserve | All blanks or 0 |
|  | IsReserveFinanced | All blanks or 0 |
|  | InitialInterestRate | 0  From 3.45 to 39.99 |
| \* | RiskScore1 | Blank, 0 to 24 |
|  | LenderAdminFee | Blank, 0, 795, 894 |
|  | DealerReserveHoldback | All blanks or 0 |
| PK | ID |  |
|  | IsEscalation | Blank, 0, -1 |
|  | DocumentationFees | All blanks or 0 |
|  | WorksheetType | Blank 0 or 1 |
|  | TValueYearLength | Blank 0 or 360 |
|  | Discount | All blank |
| ? | VarianceRate | Values from -27.99 to +14, 1 row set to -421 |
| \* | PriceToRiskRate | See also RiskTier  [blank]  18.99  20.99  24.99  27.99 |
|  | IsReserveFromPromotion | All blank or 0 |
|  | IsManuallyReserve | All blank |
|  | GSTRate | Goods and Services Tax rate |
|  | PSTRate | Provincial Sales Tax rate |
|  | BasicPayment | All blank or 0 |
|  | IsPPSAFinaced | All blank or -1 |
|  | IsResidualValueInsuranceFinaced | All blank or -1 |
|  | ResidualValueInsurance | All blank or 0 |
|  | SecurityDeposit | All blank or 0 |
|  | PromoType | All blank or 0 |
|  | PromotionDays | All blank or 0 |
|  | Channel | All blank |
|  | AdditionalChannelData | All blank |
|  | ChannelType | All blank |
|  | SalePriceSubtotal |  |
|  | ExtendedServiceContractTaxGST | GST on Extended Service |
|  | ExtendedServiceContractTaxPST | PST on Extended Service |
|  | TaxOnReplacementWarrantyGST | GST on Replacement Warranty |
|  | TaxOnReplacementWarrantyPST | PST on Replacement Warranty |
|  | InsuranceTaxGST | GST on Insurance |
|  | InsuranceTaxPST | PST on Insurance |
|  | AdminFeeGST | GST on admin fees |
|  | AdminFeePST | PST on admin fees |
|  | Installation\_Delivery\_FeeGST | GST on installation and delivery fee |
|  | Installation\_Delivery\_FeePST | PST on installation and delivery fee |
|  | OtherTaxableGST | GST on other fees |
|  | OtherTaxablePST | PST on other fees |
|  | DocumentationFeesGST | GST on documentation |
|  | DocumentationFeesPST | PST on documentation |
|  | CashSalePriceGST | GST on sale price |
|  | CashSalePricePST | PST on sale price |
|  | TradeInGST | GST on trade in |
|  | TradeInPST | PST on trade in |
|  | SpecificSubventionIndicator | All blanks |
|  | CombinedTDSR | Total Debt Service Ratio for both applicants  Ratio of debt to income (monthly) |
|  | App1TDSR | Total Debt Service Ratio for applicant  Ratio of debt to income (monthly) |
|  | App2TDSR | Total Debt Service Ratio for co-applicant  Ratio of debt to income (monthly) |
| \* | LTV1 | Loan to Value?  Why 1, 2, 3? |
| ? | LTV2 | Loan to Value?  Why 1, 2, 3? |
| ? | LTV3 | Loan to Value?  Why 1, 2, 3? |
| ? | AllInLTV | Loan to Value? |
|  | CombinedMonthlyDebts | Combined Monthly debt payment for both applicant |
|  | App1MonthlyDebts | Monthly debt payment for applicant 1 |
|  | App2MonthlyDebts | Monthly debt payment for applicant 2 |
|  | DealerStatusDecision | All blank |
| \* | TotalSubmittedIncome | Total Monthly Submitted Income for both applicants |
| \* | TotalVierifiedIncome | Total Monthly Verified Income for both applicants |
| \* | CombinedIncome | Total Monthly Income for both applicants |
| \* | TotalDebtsPayment | Total Monthly debt payments for both applicants |
|  | Applicant1Income | Monthly Income for applicant 1  Why 3: Income, Submitted and Verified. |
|  | Applicant1TotalDebtsPayment | Monthly debt payments for applicant 1 |
|  | Applicant2Income | Monthly Income for applicant 2 |
|  | Applicant2TotalDebtsPayment | Monthly debt payments for applicant |
|  | Applicant1SubmittedIncome | Monthly Submitted Income for applicant 1 |
|  | Applicant1VerifiedIncome | Monthly Verified Income for applicant 1 |
|  | Applicant2SubmittedIncome | Monthly Income for applicant 2 |
|  | Applicant2VerifiedIncome | Monthly Verified Income for applicant 2 |
|  | OtherWarranty | All blanks or 0 |
|  | PSTTaxableFee | All blanks or 0 |
|  | GSTHSTTaxableFee | All blanks or 0 |
|  | ApplicationTotalExposure | All blanks or 0 |
|  | ApplicantTotalExposure | All blanks |
|  | App1PTI | Applicant 1 Pre Tax Income?  Does not appear to be. Most values are decimals between 0 and 1. |
|  | RateBump | Blanks, 0, 1, 420 |
|  | RateBumpAmount | All blanks or 0 |
|  | TValueDateCounting | All blanks or 0 |
|  | InsuranceTaxAmount | Amount of taxes calculated on Insurance |

## ETL

Table not used in our analysis.

# Customer

PK: ID

Rows: 176,754

Duplicate Rows: None

|  |  |  |
| --- | --- | --- |
| Keep? |  |  |
| \* | ApplicationID | Internal ID for Application |
| \* | SequenceNumber | Sequence # for each application within an application.  Assuming # 1 is the main purchaser of the vehicle.  Blank, 1, 2, 3 |
|  | RelationType | Blank, 1, 2, 3 |
| \* | RelationCode | Blank  BR – Brother  CL – Common Law  GF – Girlfriend ?  GM – Grandmother ?  IL – In law  MP -  OT – Other ?  PR - Parent  SB -  SP - Spouse  SR – Sister |
|  | Prefix |  |
| \* | Gender | Blank  M – Male  F – Female |
| \* | Age |  |
| \* | MaritalStatus | CL – Common Law  DV – Divorced  MR – Married  SG - Single  SP - Separated  WD - Widowed |
| \* | ResidentType | Other  Own Free & Clear  Own Free and Clear  Own Mobile Home  Own with Mortgage  Parent  Rent  Reserve Housing |
| \* | TimeAtLocationYear | Number of years residing at current residence |
|  | TimeAtLocationMonth | Number of months residing at current residence |
| \* | AddressType | Type of address at current residence  PB – Postal Box  RR – Rural road  ST – Street |
| \* | Province | Province of residence |
| \* | PostalCode | Postal code of residence |
|  | PrevTimeAtLocationYear | Number of years residing at previous residence |
|  | PrevTimeAtLocationMonth | Number of months residing at previous residence |
|  | PrevAddressType | Type of address at previous residence  PB – Postal Box  RR – Rural road  ST – Street |
|  | PrevCity | City of previous residence |
|  | PrevProvince | Province of previous residence |
|  | PrevPostalCode | Postal Code of previous residence |
| \* | EmpType | Current type of employment  At home  Executive  Labourer  Management  Office Staff  Other  Production  Professional  Retired  Sales  Self-Employed  Service  Student  Trades  Unemployed |
| \* | EmpStatus | Current employment status  CON - Contractor  FT – Full Time  FTP – Full time permanent  Other  PTC – Part time casual  PTR – Part time R…… RET – Retired  SE – Seasonal ?  SEAS – Seasonal all summer?  SEAW – Seasonal all winter? |
|  | EmpDirection | Which way is work from home ?  E – East  N – North  NE – North East  NW – North West  O – West  S – South  SE – South East  SW – South West  W – West |
|  | EmpCity | City of current employer |
| \* | EmpProvince | Province of current employer |
| \* | EmpPostalCode | Postal Code of current employer |
| \* | EmpYears | Years working at current employer |
|  | EmpMonths | Months working at current employer |
|  | PrevEmpType | Type of previous employment  At home  Executive  Labourer  Management  Office Staff  Other  Production  Professional  Retired  Sales  Self-Employed  Service  Student  Trades  Unemployed |
|  | PrevEmpStatus | Previous employment status  FT – Full Time  FTP – Full time permanent  Other  PTC – Part time casual  PTR – Part time retired  RET – Retired  SE – Seasonal?  SEAS – Seasonal all summer?  SEAW – Seasonal all winter? |
|  | PrevOccupation | Previous Occupation.  Free text field. |
|  | PrevEmpAddressType | Type of address at previous employer  PB – Postal Box  RR – Rural road  ST – Street |
|  | PrevEmpSuiteNo | Address details of previous employer. |
|  | PrevEmpStreetNumber | Address details of previous employer. |
|  | PrevEmpStreetName | Address details of previous employer. |
|  | PrevEmpStreetType | Address details of previous employer. |
|  | PrevEmpDirection | Address details of previous employer. |
|  | PrevEmpCity | Address details of previous employer. City |
|  | PrevEmpProvince | Address details of previous employer. Province |
| \* | PrevEmpYears | Number of years working at previous employer. |
|  | PrevEmpMonths | Number of months working at previous employer. |
|  | HomeMarketValue | Market value of home |
|  | MortgageAmount | Current remaining mortgage on home |
|  | MonthlyPayment | Monthly payment for residence (mortgage or rent) |
|  | GrossIncome | Total income before taxes and deductions.  See also GrossIncomeFrq |
|  | GrossIncomeFrq | Frequency of amount specified in GrossIncome  A – Annual  M – Monthly  W – Weekly |
|  | OtherIncomeType | Category for Other source of income  Car Allowance  Child Support/Alimony  Disability Payments  Investment Income  Other  Pensions  Rental Income  Workers Compensation |
|  | Other | Amount of other source of income |
|  | OtherFrq | Frequency of other source of income |
|  | OtherDescription | Description of other source of income  Free text. |
| \* | AnnualTotal | Total annual gross income |
|  | Bnk | Which bank does applicant work with? |
| \* | AssetType1Value | Value of owned Asset # 1 |
|  | AssetType2 | Type of owned asset # 2  Blank  Auto  Other  Property  Savings  NOTE: not included for asset 1 |
|  | AssetType2Desc | Description of owned asset # 2  NOTE: not included for asset 1 |
| \* | AssetType2Value | Value of owned Asset # 2 |
|  | LibType1 | Type of Liability # 1  Credit Card  Line of Credit  Loan  Mortgage  Other |
|  | LibType1Desc | Description of Liability # 1 |
|  | LibType1MP | Monthly payment towards repayment of Liability # 1 |
|  | LibType1Bal | Balance remaining to repay Liability # 1 |
|  | LibType2 |  |
|  | LibType2Desc |  |
|  | LibType2MP |  |
|  | LibType2Bal |  |
|  | LibType3 |  |
|  | LibType3Desc |  |
|  | LibType3MP |  |
|  | LibType3Bal |  |
|  | ThirdPartyDetermination | Blank, Yes, No |
|  | ReturningCustomerFlag | Blank 0, -1 |
|  | CustomerMatch | All blank |
|  | DeclinedCustomerFlag | Blank 0, -1 |
|  | ShoppingCustomerFlag | Blank 0, -1 |
|  | OtherIncomeUsedAt | Blank, 0, 50, 100 |
|  | UseCoApplicantIncome | Was co applicant income necessary to approve loan  Blank 0, -1 |
|  | PreferredLanguage |  |
|  | ReturningCustomerMatch | All blanks |
|  | ReturningCustomerMatchEdw | All blanks |
| PK | ID | INTERNAL UNIQUE ID |
| \* | CustomerType | Blank or 1 |
|  | UpdateCreditBureau | Blank 0 or 1 |
|  | EmployeeIndicator | Blank, no, yes |
| Filter | IsDeleted | Blank, 0, -1. Only keep = 0 |
|  | ShoppingCustomerMatch | A match was identified. This customer has made one or more other loan applications.  Semicolon separated list of IDs |
|  | DefaultCreditBureau | EQ  TU |
|  | PrevOrderForAppIncome | 1  2  3 |

## ETL

Create a new table: **Customer\_NotDeleted**

Rows: It will only include customers that were not deleted.

Columns: Only columns specified in table above.

SELECT Customer.ApplicationID, Customer.SequenceNumber, Customer.RelationCode, Customer.Gender, Customer.Age, Customer.MaritalStatus, Customer.ResidentType, Customer.AddressType, Customer.TimeAtLocationYear, Customer.PostalCode, Customer.Province, Customer.EmpType, Customer.EmpStatus, Customer.EmpPostalCode, Customer.EmpProvince, Customer.EmpYears, Customer.PrevEmpYears, Customer.AnnualTotal, Customer.AssetType1Value, Customer.AssetType2Value INTO Customers\_NotDeleted

FROM Customer

WHERE (((Customer.IsDeleted)=0));

For dataset received in June 2019, that would include 174,582 rows.

# Decisions

Unable to identify a unique PK.

Rows: 580,584

Duplicate Rows: 5,096

Very similar to CreditDecisions

|  |  |  |
| --- | --- | --- |
| Keep? |  |  |
| \* | DecisionID | KEY. Not unique. |
| \* | ApplicationID | KEY. Not unique  DecisionID + ApplicationID also not unique. |
| \* | DealerCode | DT57875, See also DealerID |
|  | SubmissionDate | Date loan application was submitted. |
|  | DecisionDate | Date a decision was made on application. |
|  | TurnaroundTime | How long did it take to make decision.  In Hours? |
| \* | CreditDecision | This value also appears in CreditDecision Table. Is it the same?  AD  AP - Approved  BK - Booked  CA - Conditional Approval  CP - Credit Pending  DE - Declined  DI  DR - Document Received  EA - Expired Application  PP - Payment Pending  WD – Withdrawn  ? What is the code of Deficient Pending  See also CreditDecisionDescription |
|  | RecommendedInterestRate | See also RiskTier and PriceToRiskRate  [blank]  18.99  20.99  24.99  27.99 |
|  | Program | Via which program was application submitted? |
| \* | CashSalePrice | Amount. Purchase amount of vehicle. |
| \* | Downpayment | Amount. Down payment on vehicle purchase. |
|  | LicenseFee | Amount. |
|  | AdminFee | Amount. |
|  | Installation\_Delivery\_Fee | Amount. |
|  | GPS\_Fee | Amount.  Blank or 0 |
|  | Lender\_Fee | Amount. |
|  | Reserve | Amount. |
|  | OtherNonTaxable | Amount. |
|  | OtherTaxable | Amount. |
|  | PPSA | Amount. Personal Property Security Act ? Lien amount on vehicle? |
| \* | AmountToFinance | Amount requested to finance purchase of vehicle. |
| \* | Term | How many months is the loan for? |
| \* | PaymentFrequency | How many payments will be made each year?  4 – once every 3 months (quarterly)  12 – once every month (monthly)  24 – twice a month (bi-monthly)  26 – once every 2 weeks (bi-weekly)  52 – once very week (weekly) |
| \* | InterestRate |  |
| \* | Payment | Amount. |
| \* | Vehicle\_Make | Make of vehicle. Ex: Ford |
| \* | Vehicle\_Model | Model of vehicle. Ex: F150 |
| \* | Vehicle\_Trim | Trim of vehicle. Ex. Lariat |
| \* | Vehicle\_Year | Year of vehicle |
| \* | BookValue | Book Value of Vehicle  Some missing values |
|  | AnalystValue | Analyst’s estimated value of vehicle  Usually same as Book Value, not always |
| \* | LTV\_Value | Loan to Value?  Amount used for LTV calculation |
| \* | LTV | Loan to Value?  Ratio of Loan to the Value of Vehicle |
| \* | SubmittedIncome | Monthly income ? |
|  | PTI | Amount. Annual Pre Tax Income?  Submitted Income \* 12 appears right. Many rows have incorrect values. |
| \* | TotalCalculatedDebt | Amount. Total debt at time of application. |
| \* | TDSR | Total Debt Service Ratio?  Ratio of debt to income (monthly) |
| FILTER | IsLastDecision | Flag. Set to 1 or blank  Only keep = 1 |
|  | ApprovedIncome | Flag. Income was verified? |
|  | ExtendedWarranty | Amount paid for extended warranty on vehicle.  0 if none purchased. |
|  | OptionalLifeInsurance | Amount paid for Life Insurance.  0 if none purchased |
|  | OptionalAHInsurance | Amount paid for Accidental Health insurance.  0 if none purchased |
|  | CreditDecisionDescription | See also CreditDecision |
| \* | RiskTier | See also PriceToRiskRate and RecommendedInterestRate  S1 – 27.99  S2 – 24.99  S3 – 20.99  S4 – 18.99  I understood Risk Tier was linked to range of AmountToFinance? But I am unable to do min, max because of bad data in AmountToFinance |
| \* | TradeInAllowance | Amount of Trade-In  0 if no trade in |
|  | Lien | Amount of lien  0 if no lien |
| \* | CurrentKms | Current kilometers on vehicle |
|  | AddsValue | Amount of value added for low km? |
|  | FirstPaymentDate | Date 1st payment is due |
|  | LenderAdminFee | Amount for Administration Fee  0 if none |
|  | DocumentationFees | Amount for Documentation Fee  0 if none (they are all 0’s) |
|  | Tax | Provincial Sales Tax rate |
|  | InsuranceTax | Tax on Insurance |
|  | ExtendedServiceContractTax | Tax on Extended Warranty |
|  | FinalTax | Amount. Sum of all taxes. |
|  | TaxOnReplacementWarranty | Tax on Warranty |
| \* | DealerID | See also DealerCode |
|  | IsFirstDecision | Flag |
|  | IsFirstManualDecision | Flag |
|  | IsLastManualDecision | Flag |
|  | IsFirstCreditDecision | Flag |
|  | IsLastCreditDecision | Flag |
|  | IsOutgoingIsLastCreditDecisionNoCP | Flag. No Credit Pending. |
|  | IsFirstCreditDecisionNoCP | Flag. No Credit Pending. |
|  | IsLastManualDecisionNoCP | Flag. No Credit Pending. |
|  | IsFirstManualDecisionNoCP | Flag. No Credit Pending. |
|  | PriceToRiskRate | See also RiskTier and RecommendedInterestRate  [blank]  18.99  20.99  24.99  27.99 |
|  | MaxiumAllowanceTerm |  |
|  | SyncDate |  |

## ETL

Create a new table: **Decisions\_Booked\_isLastDecision\_Distinct**

Rows: It will only include distinct rows that were last decisions that resulted in a booked loan.

Columns: Only columns specified in table above.

SELECT DISTINCT Decisions.DecisionID, Decisions.ApplicationID, Decisions.DealerCode, Decisions.DealerID, Decisions.Vehicle\_Make, Decisions.Vehicle\_Model, Decisions.Vehicle\_Trim, Decisions.Vehicle\_Year, Decisions.CurrentKms, Decisions.CreditDecision, Decisions.CashSalePrice, Decisions.Downpayment, Decisions.AmountToFinance, Decisions.Term, Decisions.PaymentFrequency, Decisions.InterestRate, Decisions.Payment, Decisions.BookValue, Decisions.LTV\_Value, Decisions.LTV, Decisions.TotalCalculatedDebt, Decisions.TDSR, Decisions.RiskTier, Decisions.TradeInAllowance INTO Decisions\_Booked\_isLastDecision\_distinct

FROM Decisions

WHERE (((Decisions.CreditDecision)="BK") AND ((Decisions.IsLastDecision)=1));

For dataset received in June 2019, that would include 5,309 rows

# IncomeData

PK: ID

Rows: 350,863

Duplicate Rows: 5

Bad Rows: ???

Conversion Errors: 635

|  |  |  |
| --- | --- | --- |
| Keep? |  |  |
| PK | ID |  |
|  | ApplicationID | For which application ID |
|  | Name | Applicant Income  Applicant Other Income  Co-App Income  Co-App Other Income  Guarantor 1 Income  Guarantor 1 Other Income |
|  | Description | Free Text - carefull  Contains returns and messes up CSV import |
|  | SubmittedIncome | Amount |
|  | AdjustedIncome | Modified amount |
|  | IsAdjusted | Flag if amount was modified  Yes  No |

## ETL

Table not required for our analysis.

# LossDetails

PK: AccountNumber + DispostionReason

Rows: 1,396

No Duplicate Rows

|  |  |  |
| --- | --- | --- |
| Keep? |  |  |
| PK | AccountNumber | 2 duplicates  23754  43074 |
|  | DispositionDate | Date of the disposition |
|  | PrincipalOwingAmount | Amount left to repay on the load |
|  | AmountsRecoveredOnDisposition | Amount recovered by disposition |
| PL | DispostionReason | EOT – End of Term  EPO -  INS – Insurance  REPO – repossession  UNWOUND  WRITE-OFF - |
|  | LoanNumber |  |

## ETL

TBD. We may require this table, but currently using another file called DataOnLosses.

# Losses

Table not required for our analysis.

# Vehicles

PK: ID

Rows: 160,881

No Duplicate Rows.

Almost self-explanatory. Will define later.

|  |  |  |
| --- | --- | --- |
| Keep? |  |  |
| \* | PortalApplicationID |  |
| \* | CreatedOn |  |
| \* | ApplicationNumber |  |
|  | ReqType |  |
| \* | DealerID |  |
| \* | ProductMake |  |
| \* | ProductModel |  |
| \* | ProductSeries |  |
|  | ProductBodyStyle |  |
| \* | ProductYear |  |
|  | ProductCondition |  |
| \* | ProductMileage |  |
|  | ProductWholesaleClean |  |
|  | ProductIncludes |  |
|  | ProductAdds |  |
|  | ProductAddsValue |  |
|  | ProductDeductsValue |  |
|  | CreditWorkflowState |  |
|  | AuditFundingWorkflowState |  |
|  | IncomeWorkflowState |  |
| \* | LenderBookValue |  |
|  | VerifiedMonthlyIncome |  |
|  | IncomeStatus |  |
| \* | MonthlyIncome |  |
|  | SubmittedIncome |  |
|  | DecStatusLast | = BK (These are the vehicles that are booked for a loan, the only ones we are interested in) |
| \* | RiskTierCode |  |
| \* | RiskScore1 |  |
|  | OriginalBookType |  |
|  | DealerName |  |
| \* | CustomerProvince |  |
|  | ReturningCustomerFlag |  |
| \* | InitialNumberOfApplicants |  |
| \* | CreditStatusLast |  |
| \* | RiskTierName |  |
| \* | LastRiskTierName |  |
| \* | LastAmountToFinance |  |
| \* | App1RiskScore |  |
| \* | App2RiskScore |  |
| \* | DealerRegion |  |
|  | CompassID |  |
| \* | CoApplicantFlag |  |
|  | LenderBookValueWithoutAddsOrDeducts |  |
| PK | ID |  |
|  | AddsMunisDeducts |  |
|  | AssetTypeID |  |
|  | CompassUniqueId |  |

## ETL

Create a new table: **Vehicles\_Booked**

Rows: It will only include vehicles that are now booked for a loan.

Columns: Only columns specified in table above.

SELECT Vehicles.PortalApplicationID, Vehicles.CreatedOn, Vehicles.ApplicationNumber, Vehicles.DealerID, Vehicles.ProductMake, Vehicles.ProductModel, Vehicles.ProductSeries, Vehicles.ProductYear, Vehicles.ProductMileage, Vehicles.LenderBookValue, Vehicles.MonthlyIncome, Vehicles.RiskTierCode, Vehicles.RiskScore1, Vehicles.CustomerProvince, Vehicles.InitialNumberOfApplicants, Vehicles.CreditStatusLast, Vehicles.RiskTierName, Vehicles.LastRiskTierName, Vehicles.LastAmountToFinance, Vehicles.App1RiskScore, Vehicles.App2RiskScore, Vehicles.DealerRegion, Vehicles.CoApplicantFlag INTO Vehicles\_Booked

FROM Vehicles

WHERE (((Vehicles.DecStatusLast)="BK"));

For dataset received in June 2019, that would include 5,320 rows.

# Data on Losses - Quarterly LPC spreadsheets

Defaulted Loan is entered in appropriate quarterly spreadsheet based on Disposition Date.

|  |  |  |
| --- | --- | --- |
| Keep? |  |  |
| \* | DT # | Dealer Track # |
| \* | Dlr | Selling Dealer name |
|  | Booked | Date deal was booked in house |
|  | Financed | Amount paid to the Dealer |
|  | Unit | Year, Make, & model of vehicle |
| \* | Contract | Contract number |
| \* | Acct# | Customer number |
| \* | Loan Amount | Financed Loan amount |
|  | Disposition Date | Date loan closed out in Loan Management System |
| \* | Disposition Reason | Disposition Reason |
| \* | Principal amount owing | Principal balance as of the last payment received on the loan prior to Disposition |
|  | Amounts recovered on Disposition | Funds received for the closeout of the loan |
| \* | Loss on Dispostion | Formula = (principal balance - amounts recovered on Disposition) |

## Summary of Issues

2019 Q1 Disposition Date conversion (Set column format to date)

2018 Q4 has only DT#, Dlr, Booked, Finance, Unit. Missing everything else.

2018 Q3 – No Issues.

2018 Q2 Missing DT#. Disposition Date conversion (Set column format to date)

2018 Q1 missing spreadsheet?

2016 Q2 to 2017 Q4 – Missing DT#, Disposition Date conversion (mix of formats)

## ETL

1. Manual step in Excel. Save each worksheet as a separate CSV file.

Trying to import directly into MS-Access from MS-Excel worksheet produces conversion errors on the Booked data field.

2016 Q2 to 2017 Q4 Disposition Date format is a mix of dd-mmm-yy, mmm dd/yy

Will require some coding to convert properly. See VBA function dispositionDate() in Appendix A

1. Manual step in MSAccess. Import CSV file into separate tables in MSAccess.

Should not receive any import or conversion errors. Verify all date fields are correct.

1. Using Append Queries. Merge All quarterly DataOnLosses into 1 master table.

Not all quarterly tables have all fields. See results of merge in following table.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Total Rows | Actual Rows Appended | Reason |
| Total |  | 646 |  |
| 2019 Q1 | 242 | 242 |  |
| 2018 Q4 | 207 | 206 | PK violation  DT19795017SON02 2018/6/14 |
| 2018 Q3 | 198 | 198 |  |
| 2018 Q2 | 224 | 0 | No DT# to link! |
| 2018 Q1 | 0 | 0 | File not provided |
| 2016 Q2 to  2017 Q4 | 528 | 0 | No DT# to link! |

# Data on Losses - Quarterly Additional recoveries

|  |  |  |  |
| --- | --- | --- | --- |
| PK? | DT # | Dealer Track # |  |
|  | Dlr | Selling Dealer name |  |
|  | Booked | Date deal was booked in house |  |
|  | Financed | Amount paid to the Dealer |  |
|  | Unit | Year, Make, & model of vehicle |  |
| \* | Contract | Contract number |  |
| \* | Acct# | Customer number |  |
| \* | Loan Amount | Financed Loan amount |  |
|  | Disposition Date | Date loan closed out in Loan Management System |  |
| \* | Disposition Reason | Disposition Reason |  |
|  | Principal amount owing | Principal balance as of the last payment received on the loan prior to Disposition |  |
|  | Amounts recovered on Disposition | Funds received for the closeout of the loan |  |
|  | Loss on Dispostion | Formula = (principal balance - amounts recovered on Disposition) |  |

## ETL

Table not used for our analysis.

# Appendix A –DispositionDate format fix

VBA FUNCTION to convert DispositionDate

Public Function dispositionDate(value As String) As Date

Rem 2 formats found in this column

Rem MAR 14/18

Rem 9-Feb-17

Rem which format is this value

Dim v1, sday, smonth, syear, sFixedDate As String

Dim PosSlash, PosSpace, PosDash1, PosDash2 As Integer

Dim day, month, year As Integer

Dim dv As Date

On Error GoTo BadDate

v1 = Trim(value)

dv = DateValue(v1)

Rem No Error so it is already a valid date

dispositionDate = dv

Exit Function

Rem ERROR - It is not a valid date

BadDate:

PosSlash = InStr(v1, "/")

If PosSlash > 0 Then

Rem Format likely is MAR 14/18

Rem fix the year

syear = Mid(v1, PosSlash + 1)

year = Int(syear)

If Len(syear) = 2 Then

year = 2000 + year

syear = Str(year)

End If

Rem find the space, extract day and month

PosSpace = InStr(v1, " ")

If PosSpace = 0 Then

Do While Not IsNumeric(Mid(v1, PosSpace + 1, 1))

PosSpace = PosSpace + 1

Loop

smonth = Left(v1, PosSpace)

Else

smonth = Left(v1, PosSpace - 1)

End If

sday = Mid(v1, PosSpace + 1, PosSlash - PosSpace - 1)

day = Int(sday)

Rem convert to proper format and to a date value

sFixedDate = sday + "-" + smonth + "-" + syear

dispositionDate = DateValue(sFixedDate)

Else

Rem 9-Feb-17

PosDash1 = InStr(v1, "-")

PosDash2 = InStr(PosDash1 + 1, v1, "-")

sFixedDate = Left(v1, PosDash2) + "20" + Mid(v1, PosDash2)

dispositionDate = DateValue(sFixedDate)

End If

End Function

# Appendix B – Creating clean dataset

The following step are to create a cleaned and complete data set. To have one complete dataset we must link each Customer row to a row in the Defaulted Loans tables.

The dataset provided does not include a direct link between

Customer -X> Vehicle -> DataOnLosses

Does Vehicle link with Customer? NO.

Customer.ApplicationID: Number

Table: Vehicles

|  |  |  |
| --- | --- | --- |
| PortalApplicationID: | String | No matches |
| ApplicationNumber: | String | No matches |
| DealerID: | number | Coincidental numerical matches. Wrong! |
| CompassID: | String | No matches |
| ID: | Number | Coincidental numerical matches. Wrong! |
| AssetTypeID: | Number | No Matches |
| CompassUniqueID: | string | No Matches. |

As per the ER Diagram shown at the beginning of this document, we must first join Customers to Decisions, Decisions to Vehicles and finally Vehicles to DataOnLosses.

The final dataset will have 1 row per customer with associated co-applicant, decision, vehicle and default information.

The sequence of the first four steps is not important, since we are generating a cleaner subset of each table.

1. DataOnLosses. See manual ETL steps for that table.

**646 rows**

1. See ETL steps for Vehicles table. Only keep rows Booked.

New table **Vehicles\_Booked** with **5,320 rows** (down from 160,881)

1. See ETL steps for Decisions table. Only keep distinct rows isLastDecision and Booked.

New table **Decisions\_Booked\_isLastDecions\_Distinct** with **5,309 rows** (down from 580,584)

1. See ETL steps for Customers table. Only keep non-deleted.

New table: **Customer\_NotDeleted** with **174,582 rows** (down from 176,754)

In the next section, we are linking those four tables together. The process starts from the Vehicle table, since this in essence is our list of booked loans.

1. Link Vehicles to DataOnLosses

Outer Join, All rows from Vehicles\_Booked linked to DataOnLosses if it exists.

New temporary table: Vehicles\_Booked\_Losses

SELECT Vehicles\_Booked.\*, [DataOnLosses LPC].\* INTO Vehicles\_Booked\_Losses

FROM Vehicles\_Booked LEFT JOIN [DataOnLosses LPC] ON Vehicles\_Booked.ApplicationNumber = [DataOnLosses LPC].[﻿DT #];

Should result in same number of rows as Vehicles\_Booked. 5,320 Rows.

1. Add Link to Decisions.

Inner Join. Match rows from Decisions\_Booked\_isLastDecision\_Distinct to Vehicles\_Booked\_losses

New temporary table: Decisions\_Vehicles\_Losses

SELECT Decisions\_Booked\_isLastDecision\_Distinct.\*, Vehicles\_Booked\_Losses.\* INTO Decisions\_Vehicles\_Losses

FROM Vehicles\_Booked\_Losses INNER JOIN Decisions\_Booked\_isLastDecision\_Distinct ON (Vehicles\_Booked\_Losses.ProductMileage = Decisions\_Booked\_isLastDecision\_Distinct.CurrentKms) AND (Vehicles\_Booked\_Losses.ProductYear = Decisions\_Booked\_isLastDecision\_Distinct.Vehicle\_Year) AND (Vehicles\_Booked\_Losses.ProductModel = Decisions\_Booked\_isLastDecision\_Distinct.Vehicle\_Model) AND (Vehicles\_Booked\_Losses.ProductMake = Decisions\_Booked\_isLastDecision\_Distinct.Vehicle\_Make) AND (Vehicles\_Booked\_Losses.DealerID = Decisions\_Booked\_isLastDecision\_Distinct.DealerID);

Should result in same number of rows as Vehicles\_Booked. 5,320 Rows.

But I only get 5305 rows. Find out why I lost 15 rows

1. Split Customer into Applicants and Co-Applicant

New table: Customers\_Applicant. Customers with Sequence\_Number = 1

SELECT Customer\_NotDeleted.\* INTO Customers\_CoApplicant

FROM Customer\_NotDeleted

WHERE (((Customer\_NotDeleted.SequenceNumber)=1));

160,448 Rows.

New table: Customers\_CoApplicant. Customers with Sequence\_Number = 2

SELECT Customer\_NotDeleted.\* INTO Customers\_CoApplicant

FROM Customer\_NotDeleted

WHERE (((Customer\_NotDeleted.SequenceNumber)=2));

14,108 Rows.

1. Link Applicant, Co-Applicant to Decisions\_Vehicles\_Losses

Outer Join:

New table: Final\_Dataset

SELECT Customers\_Applicant.\*, Customers\_CoApplicant.\*, Decisions\_Vehicles\_Losses.\* INTO Final\_Dataset

FROM Customers\_CoApplicant RIGHT JOIN (Customers\_Applicant RIGHT JOIN Decisions\_Vehicles\_Losses ON Customers\_Applicant.ApplicationID = Decisions\_Vehicles\_Losses.ApplicationID) ON Customers\_CoApplicant.ApplicationID = Decisions\_Vehicles\_Losses.ApplicationID;

Should result in same number of rows as Decisions\_Vehicles\_Losses. 5,305 Rows. Will be 5320 when step 6 is fixed.