UCD MSc Business Analytics

Capstone Project:

Predicting Transactions Times

# Iteration 2 - Data Understanding

Eoin Carroll

Kieron Ellis

Draft Rev 2 15/6/17 – Variables included in modelled dataset highlighted in green. Note on TimeTaken

Document purpose: This is a working document intended to capture our understanding of the available data. This will eventually form part of the final project report.

Latest dataset: Dataset from Cosmic launch (6th Feb) to End March. Pulled on 20/4/17 and provided to us on 26/4/17.

Filename: UCD\_Data\_20170420\_1.xlsx

To predict the processing times for cases, we have been provided with historic data. The dataset that we are working from was extracted from the newly implemented COSMIC records logging system. The data is in the form of an excel workbook with 4 data sheets along with some additional sheets which include how the data was obtained – ie. what query was used.

1. vw\_Incident – 13018 items
2. vw\_HoldActivity – 9104 items
3. vw\_AuditHistory – 16148 items
4. vw\_PackageTriageEntry – 9075 items

Every case includes a different number of transactions but we are only interested in the time to process a case. The incident worksheet contains a list of cases. HoldActivity, AuditHistory and PackageTriageEntry contain transaction level details. The PackageTriageEntry form has been disabled and will not be available in the future unless we determine that there is value in requesting it.

There is a seasonality aspect to the data – iteration 3.

Merging and prediction transaction level details will be in iteration 3.

The following tables lay out the variables in each sheet, their significance, our pre-processing decision and the reasoning. The decisions at this point will be reviewed at each iteration and could be changed. Note that red means that variable needs extra work.

The decisions were implemented in a pre-processing program and a cleaned dataset was created. Modelling was then trialled on the incident data. See modelling document.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Worksheet** | **Variable Name** | **Variable Use** | **Pre-Processing** | **Reasoning** |
| Incident | TicketNumber | Identification of case | Keep as is | Need for case ID |
| Incident | IncidentId | System identifier for the case | Delete column | Use TicketNumber for ID |
| Incident | Created\_On | Created in COSMIC by BPO team | Use to calculate time taken – then delete | Needed for time taken only |
| Incident | Receiveddate | Operations Centre received the pack | Delete column | Manual entry – not reliable. Using other fields for time |
| Incident | ResolvedDate | Date the case was resolved | Use to calculate time taken – then delete | Needed for time taken only |
| Incident | Queue | The location of case at time of pulling data | Combine types, One hot encoding | Too many options for direct one hot |
| Incident | StatusReason | Outcome of case | Filter out "Rejected", One hot | Not interested in rejected |
| Incident | Priority |  | Mapping | Ordinal variable |
| Incident | ValidCase |  | Filter by 1, Delete column (min\_variable\_types function will clear this) | Only interested in 1 entries |
| Incident | BusinessFunction |  | Delete column (min\_variable\_types, function will clear this) | Only 1 variable type |
| Incident | LineOfBusiness |  | Delete column (min\_variable\_types, function will clear this) | Only 1 variable type |
| Incident | Program |  | Filter by "Enterprise", Delete column (min\_variable\_types function will clear this) | Only want "Enterprise" entries |
| Incident | CaseType |  | Delete column (min\_variable\_types, function will clear this) | Only 1 variable type |
| Incident | CaseSubTypes |  | Delete column (min\_variable\_types, function will clear this) | Only 1 variable type |
| Incident | Reason |  | Delete column (min\_variable\_types, function will clear this) | Only 1 variable type |
| Incident | SubReason |  | One hot |  |
| Incident | SubSubReason |  | Delete column (min\_entries function will clear this) | No entries present |
| Incident | CountrySource | Where did case come from | Combine types, One hot encoding | Important but need to place in buckets |
| Incident | CountryProcessed | Where is case being processed | Combine types, One hot encoding | Important but need to place in buckets |
| Incident | SalesLocation | Where is sale being made to | Combine types, One hot encoding | Important but need to place in buckets |
| Incident | ROCName | Regional operating centre name | One hot encoding |  |
| Incident | CaseRevenue |  | Keep as is but normalise | Do not know what this does |
| Incident | CurrencyName |  | Delete column | We have all in USD, can delete others |
| Incident | IsoCurrencyCode |  | Delete column | We have all in USD, can delete others |
| Incident | RevenueImpactAmount |  | Delete column | Mostly NULL entries |
| Incident | IsGovernment |  | Keep as is |  |
| Incident | IsAudited |  | Keep as is | Mostly 0 entries |
| Incident | Language |  | Delete column, (min\_variable\_types function will clear this) | Only 1 variable type (after filtering) |
| Incident | LanguageName |  | Filter by "English", Delete column (min\_variable\_types function will clear this) | Only want "English" entries |
| Incident | sourcesystem |  | One hot encoding |  |
| Incident | caseOriginCode |  | Delete column | Do not know what this does |
| Incident | Source |  | One hot encoding |  |
| Incident | pendingemails |  | Delete column | Mostly 0 entries |
| Incident | RelatedCases |  | Delete column | Don’t know how else to include this |
| Incident | Auditresult |  | Delete column | Mostly 0 entries |
| Incident | PendingRevenue |  | Delete column | Mostly 0 entries |
| Incident | Requestspercase |  | Delete column | Mostly 1 entries |
| Incident | Totalbillabletime |  | Delete column | Mostly 0 entries |
| Incident | Totaltime |  | Delete column | Mostly 0 entries |
| Incident | CreditAmount |  | Delete column | Mostly NULL entries |
| Incident | DebitAmount |  | Delete column | Mostly NULL entries |
| Incident | OrderAmount |  | Delete column | Not enough entries |
| Incident | InvoiceAmount |  | Delete column | Mostly NULL entries |
| Incident | Deleted |  | Delete column | No entries |
| Incident | TotalIdleTime |  | Delete column | Not useful |
| Incident | TotalWaitTime |  | Delete column | Not useful |
| Incident | WorkbenchGroup |  | Delete column | Do not know what this does |
| Incident | Workbench |  | Delete for now (then One hot in groups) | Do not know what this does |
| Incident | RejectionReason |  | Delete column | Mostly NULL entries |
| Incident | RejectionSubReason |  | Delete column | Mostly NULL entries |
| Incident | PackageNumber |  | Delete column | Mostly NULL entries |
| Incident | RequiredThreshold |  | Delete column | Mostly 0 entries |
| Incident | Slipped |  | Delete column | Mostly 0 entries |
| Incident | DefectiveCase |  | Delete column | Mostly 0 entries |
| Incident | OLSRevenue |  | Keep as is |  |
| Incident | Revenutype |  | One hot encoding |  |
| Incident | ProcessName |  | Delete column | Only 1 variable type |
| Incident | StageName |  | One hot encoding (map next time) | Ordinal variable |
| Incident | AmountinUSD |  | Keep as is, but normalise | Important variable |
| Incident | IsMagnumCase |  | Keep as is |  |
| Incident | IsSignature |  | Keep as is |  |
| Incident | Complexity |  | Mapping | Ordinal variable |
| Incident | Numberofreactivations |  | Delete column | not enough entries |
| Incident | NumberofChildIncidents |  | Delete column | Mostly NULL entries |
| Incident | ParentCase |  | Delete column | Mostly NULL entries |
| Incident | Referencesystem |  | Delete column | Not enough entries |
| Incident | StateCode |  | Delete column | Only 1 variable type |
| Incident | Isrevenueimpacting |  | Delete column | Mostly 0 entries |
| Incident | IsSOXCase | Binary: A legal document requirement | Keep, fill NULLS with 0s | Important variable. Assume 0 or Null means not SOX |

Note that Y variable “TimeTaken” was calculated from created on and resolved on columns.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Worksheet** | **Variable Use** | **Variable Name** | **Pre-Processing** | **Reasoning** |
| HoldActivity | Links to case ID | TicketNumber | Keep as is | Need for case ID |
| HoldActivity | System ID | ActivityId | Delete column | Use TicketNumber for ID |
| HoldActivity |  | RegardingObjectId | Keep as is | Need for ID |
| HoldActivity |  | HoldDuration | Keep as is |  |
| HoldActivity |  | Statuscode | Filter by “Completed”, Delete column | Only interested in completed |
| HoldActivity |  | Statecode | Delete column | Duplicate of Statuscode |
| HoldActivity |  | TimeZoneRuleVersionNumber | Keep as is |  |
| HoldActivity |  | HoldTypeName | Filter out "Internal", one hot | Not interested in Internal |
| HoldActivity |  | Reason | One hot encoding |  |
| HoldActivity |  | AssignedToGroup | One hot encoding |  |
| HoldActivity |  | StartTime | Calculate time, delete |  |
| HoldActivity |  | EndTime | Calculate time, delete |  |
| HoldActivity |  | IsThreeStrikeRule | Delete column | Only 1 variable type |
| HoldActivity |  | HoldSubReason | Delete column | Mostly nulls |
| HoldActivity |  | Modified\_On | Delete column | Use HoldDuration for time |
| HoldActivity |  | Deleted | Delete column | No entries |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Worksheet** | **Variable Use** | **Variable Name** | **Pre-Processing** | **Reasoning** |
| AuditHistory | Links to case ID | TicketNumber | Keep as is | Need for case ID |
| AuditHistory | System ID | AuditHistoryId | Delete column | Use TicketNumber for ID |
| AuditHistory |  | AuditId | Keep as is | Need for ID |
| AuditHistory |  | EntityId | Keep as is | Need for ID |
| AuditHistory |  | EntityLogicalName | Delete column (min\_variable\_types function will clear this) | Only 1 variable type |
| AuditHistory |  | CaseId | Keep as is | Need for ID |
| AuditHistory |  | NewValue | Combine (NewValue, OldValue, Action) | Movement through system |
| AuditHistory |  | OldValue | Combine (NewValue, OldValue, Action) |  |
| AuditHistory |  | Action | Combine (NewValue, OldValue, Action) |  |
| AuditHistory |  | Attribute | Delete column (min\_variable\_types function will clear this) | Only 1 variable type |
| AuditHistory |  | TimeStamp | Delete column | Use Created\_on |
| AuditHistory |  | Deleted | Delete column (min\_entries function will clear this) | No entries |
| AuditHistory |  | TenantId | Delete column (min\_variable\_types function will clear this) | Only 1 variable type |
| AuditHistory |  | VersionNumber | Delete column (min\_variable\_types function will clear this) | Only 1 variable type |
| AuditHistory |  | StateCode | Delete column (min\_variable\_types function will clear this) | Only 1 variable type |
| AuditHistory |  | Status | Delete column (min\_variable\_types function will clear this) | Only 1 variable type |
| AuditHistory |  | StatusCode | Delete column (min\_variable\_types function will clear this) | Only 1 variable type |
| AuditHistory |  | StatusReason | Delete column (min\_variable\_types function will clear this) | Only 1 variable type |
| AuditHistory |  | UTCConversionTimeZoneCode | Keep as is |  |
| AuditHistory |  | Created\_On | Use for time, Delete column |  |
| AuditHistory |  | Modified\_On | Use for time, Delete column |  |
| AuditHistory |  | ImportSequenceNumber | Delete column (min\_entries function will clear this) | No entries |
| AuditHistory |  | OverriddenCreatedOn | Delete column (min\_entries function will clear this) | No entries |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Worksheet** | **Variable Use** | **Variable Name** | **Pre-Processing** | **Reasoning** |
| PackageTriageEntry | Links to case ID | TicketNumber | Keep as is | Need for case ID |
| PackageTriageEntry | System ID | PackageTriageEntryId | Delete column | Use TicketNumber for ID |
| PackageTriageEntry |  | EntryType | One hot encoding |  |
| PackageTriageEntry |  | EntryLevel | One hot encoding |  |
| PackageTriageEntry |  | EntryProcess | One hot encoding |  |
| PackageTriageEntry |  | PCNStatus | Delete column | Mostly 0s |
| PackageTriageEntry |  | SAPStatus | Delete column | Mostly 0s |
| PackageTriageEntry |  | PackageTriageFormId | Keep as is | Need for ID |
| PackageTriageEntry |  | StateCode | Delete column | Only 1 variable type |
| PackageTriageEntry |  | StatusCode | Delete column | Only 1 variable type |
| PackageTriageEntry |  | Created\_On | Use for time, Delete column |  |
| PackageTriageEntry |  | Modified\_On | Use for time, Delete column |  |