|  |
| --- |
| VTS Data Extraction |
| Data Extraction |

Revision Control

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Version No | Date | Change / Activity | By | Approver | Circulation |
| PA1 | 31-Dec-18 | Initial Draft | Saurabh | Rajeev | Internal |
| A |  |  |  |  |  |

Reference Documents

|  |  |  |  |
| --- | --- | --- | --- |
| S No | Document Name | Version / Date | Owner |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Copyright: SENSORISE 2018

SENSORISE Confidential and Proprietary Information

All Rights Reserved

The contents of this document are restricted for circulation; please use pursuant to company instructions and the terms of the Non-Disclosure Agreement.

# 1 Introduction

* This document explains the extraction protocols followed to get set of points extracted from original port file with respect to each and every day of data
* Raw data available to us contains five port files namely ( in order of size from largest to smallest):
* 8092
* 8091
* 8090
* 8094
* 8099
* The data in these port files contains various fields related to the data, which is basically a VTS stamp displaying vehicle properties (to be discussed later the document)
* Data from the running vehicles is captured by multiple data capture servers which are listing on different socket ports. One port is assigned to 1 or more OEMs.
* Since for driver behavior analysis data needs to be segregated vehicle wise a standard protocol has to be defined to get vehicle information individually
* For this purpose, it is assumed each and every vehicle has a unique identifier associated with it, i.e IMEI number which shows vehicle information categorically
* Each of these port files have data from date 1 July 2018 to 31 July 2018 in .csv format which comprises VTS stamps several number of identifiers(IMEI)
* Various fields need to be extracted for each and every IMEI and structured so that data can be preprocessed for further analysis

# 2. Data Format

1. Each data file consists of data received in a given date and a given port
2. File is semicolon (;) separated
3. One file consists of several packets and each packet consists of 1 or more data records.
4. Packet represents data received from the device in one session transmission. Normally each packet consists of 1 data record with current location and device status. In case device was not able to send the packet, it is kept in the buffer and sent in the next transmission as History record.
5. Each packet has the following structure:
   1. Fields are semicolon separated
   2. Fields:
      1. Recordid: This is a unique id for each packet across all the files and days
      2. Datetimestamp: Date and time when this packet was received
      3. Datamessage: message received from the device. This contains 1 or more location and device status records.
      4. Clientfdn: This gives server and client socket details.
      5. Clientport: Port number of the server listing for the data
6. Each Record in the packet has the following Structure
   1. Starts with $ and ends with \*, rest of the format is given below.
   2. Multiple records will follow each other as follows

$......\*$.......\*$......\*

1. Each Record is comma (,) delimited.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| # | Field | Description | Validation | RAW Data |
| 1 | **Start Character** | $ | Yes | $ |
| 2 | **Header** | The header of the packet/ identifier | Yes | rsm |
| 3 | **Vendor ID** | Vendor identification header | No | ICAT0260a |
| 4 | **Firmware Version** | Version details of the Firmware used in EX.1.0.0 | No | 1.0.0 |
| 5 | **Packet Type** | Specify the packet type – | Yes | **NR** |
| NR = Normal |
| EA = Emergency Alert (Optional) TA = Tamper Alert |
| HP = Health Packet |
| IN = Ignition On |
| IF = Ignition Off |
| BD = Battery Disconnect BR = Battery Reconnect BL = Battery Low |
| TS = Trip Start |
| TE = Trip End |
| 6 | **Packet Status** | L=Live or H= History | Yes | L |
| 7 | **IMEI** | Identified of the sending unit. 15 digit standard unique IMEI no. | Yes | 868324020300260 |
| 8 | **Vehicle Reg. No** | Mapped vehicle registration number | No | DL0260a |
| 9 | **GPS Fix** | 1 = GPS fix OR 0 = GPS invalid | Yes | 1 |
| 10 | **Date** | Date value as per GPS date time (ddmmyy) | Yes (dtok)  Yymmdd / yyyymmdd /ddmmyy/ddmmyyyy | 6082016 |
| 11 | **Time** | Time value as per GPS date time in UTC format (hhmmss) | Yes HHMMSS | 135406 |
| 12 | **Latitude** | Latitude value in decimal degrees (with minimum 6 decimal places) | Yes | 28.35886 |
| 13 | **Latitude Dir.** | Latitude Direction. Example N=North, S= South | Yes  N or S | N |
| 14 | **Longitude** | Longitude value in decimal degrees (with minimum 6 decimal places) | Yes  Decimal | 76.92736 |
| 15 | **Longitude Dir.** | Longitude Direction. Example E=East, W= West | Yes  E or W | E |
| 16 | **Speed** | Speed in km/hr | Yes | 0 |
| 17 | **Heading** | Course over ground in degrees | No | 313 |
| 18 | **No of Satellites** | Number of satellites available for fix | Yes | 4 |
| 19 | **Altitude** | Altitude of the device in meters | Yes | 188.7 |
| 20 | **PDOP** | Positional dilution of precision | No | 3.03 |
| 21 | **ODO** | Total Distance Travelled in Km (from activation of unit to till last packet) | Yes | 0 |
| 22 | **Network Operator Name** | Name of Network Operator. | No | airtel |
| 23 | **Ignition** | 1= Ign On , 0 = Ign Off | Yes | 0 |
| 24 | **Main Power Status** | 0 = Vehicle Battery Disconnected 1= Vehicle Battery Reconnected | Yes | 1 |
| 25 | **Emergency Status (Optional)** | 1= On , 0 = Off | No | 0 |
| 26 | **Tamper Alert** | C = Cover Closed , O = Cover Open | Yes | C |
| 27 | **GSM Signal Strength** | Value Ranging from 0 – 31 | Yes | 26 |
| 28 | **MCC** | Mobile Country Code | Yes | 404 |
| 29 | **MNC** | Mobile Network Code | Yes | 10 |
| 30 | **LAC** | Location Area Code | No | 209 |
| 31 | **Cell ID** | GSM Cell ID | No | 6d17 |
| 32 | **LAC1** |  | No |  |
| 33 | **NMR2** |  | No |  |
| 34 | **LAC2** |  | No |  |
| 35 | **NMR3** |  | No |  |
| 36 | **LAC3** |  | No |  |
| 37 | **NMR4** |  | No |  |
| 38 | **LAC4** |  | No |  |
| 39 | **Digital Input Status** | 4 external digital input status (Status of Input 1 to Input 3 (0=Off; 1=On)) | Yes | 0,0,0,0 or 0000 or 1001 |
| 40 | **Digital Output Status** | 2 external digital output status (0=Off; 1=On) | Yes | 1,1 |
| 41 | **Frame Number** | Sequence Number of the messages (000001 to 999999) |  | 65917 |
| 42 | **Checksum** | Insures No error in transmission (optional) | No Ignore |  |
| 43 | **End Character** | Indicated End of the frame | No | \* |

Validation data  
Data should be validated with the following fields else rejected:

|  |  |  |
| --- | --- | --- |
| S.No. | Field | Criterion |
| 1 | Number of Records | The number of records in each filed message ( record) should be more than 30 |
| 2 | Start Character | The start character ($) should be found on each packet |
| 3 | Packet Type | Following packet type should be present:   1. NR: Normal 2. EA: Emergency Alert 3. HP: Health Packet 4. IN,IF: Ignition On and Off 5. TA: Temper Alert 6. BD, BR: Battery Disconnected/ Reconnected 7. BL: Battery Low 8. TS, TE: Trip Start and Trip End 9. HA,HB,RT: Harsh Acceleration, Break, Rash Turn |
| 4 | Packet Status | L: Live  H: History |
| 5 | IMEI | IMEI length should be 15 and should not start with 0 |
| 6 | Latitude | It should be a float value |
| 7 | Longitude | It should be a float value |
| 8 | Latitude Direction | It should be “N” |
| 9 | Longitude Direction | It should be “S” |
| 10 | Date Time | Format to be validated |
| 11 | PDOP | It should be float value and less than 10 |
| 12 | Tamper alert | It should be “C” or “O” |

6. Function Used:

## Read and write file:

### 6.1.1**. Introduction**

* This function reads line by line each line each record of a port file and processes it for generating extracted cleaned file
* The role of this function is to give extracted records in clean format in respective IMEI, Track and error summary files

### 6.1.2**.Input**

* Port file to be processed
* Directory where files are to be stored

### 6.1.3**. Output**

1. Track file containing port name and date in .csv format which contains fields UID and Record
2. Error Summary File containing port name and date in .csv format which contains following fields:
   1. UID : Since each packet comes with an identifier it signifies it.
   2. Port : Port number from where record is obtained
   3. Record Index : Record index of the port file
   4. Description : Reason for which the record is an error
   5. Record : The whole record which was rejected

1. IMEI file which contains following fields:

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No** | **Data** | **Type** | **Comments** |
| 1 | IMEI | String | Device Identification |
| 2 | VendorID | String | Device Supplier |
| 3 | VehicleNumber | String | Vehicle Registration Number |
| 4 | MNO | String | Operator Name |
| 5 | MCC | String | Operator Country Code |
| 6 | MNC | String | Operator Code |
| 7 | LAC | String | Location Code of Cell tower |
| 8 | CellID | String | Tower ID |
| 9 | PacketType | String | Normal, Emergency, Health, Ignition.. |
| 10 | PacketStatus | String | Live or History |
| 11 | DateTime | DateTime | Packet Generated Date Time |
| 12 | GPSFix | Numeric | 1 (No Error) and 0(Error) |
| 13 | Lat | Numeric | Accuracy of 7 digits |
| 14 | LatDir | String | N/S Direction |
| 15 | Lon | Numeric | Accuracy of 7 digits |
| 16 | LonDir | String | E/W Direction |
| 17 | InstSpeed | Numeric | Instantaneous Speed |
| 18 | Heading | Numeric | Direction in degrees |
| 19 | Altitude | Numeric | Hight of the vehicle |
| 20 | PDOP | Numeric | Goodness of GPS, normally <7 and Good accuracy means < 4 |
| 21 | Ignition | Numeric | 1 (On) 0 (off) |
| 22 | MainPowerStatus | Numeric | 0 (Disconnected) 1 (Connected) |
| 23 | EmergencyStatus | Numeric | < 0 (Missing) 0 (Normal) 1 (Emergency) |
| 24 | TemperAlert | String | C (Cover Closed), O (Open) |
| 25 | GPSSignalStrength | Numeric | Can be used to decide on struct on GPS data |
| 26 | DigitalInput1 | Numeric | 0 or 1 |
| 27 | DigitalInput2 | Numeric | 0 or 1 |
| 28 | DigitalInput3 | Numeric | 0 or 1 |
| 29 | DigitalInput4 | Numeric | 0 or 1 |
| 30 | DigitalOutput1 | Numeric | 0 or 1 |
| 31 | DigitalOutput2 | Numeric | 0 or 1 |
| 32 | DigitalOutput3 | Numeric | 0 or 1 |
| 33 | DigitalOutput4 | Numeric | 0 or 1 |
| 34 | FrameNumber | Numeric | 000001 to 999999 |
|  |  |  |  |

1. File Summary containing following fields:

|  |  |  |
| --- | --- | --- |
| S.No. | Field | Comment |
| 1 | File Name | File name of the processed file |
| 2 | Packet | Total number of packet processed on file(Unique UID’s in file) |
| 3 | Record | Total records processed in packet |
| 4 | Good Record | Total number of good records which are processed in file |
| 5 | Error Record | Total number of error records which are processed in file |

## Process

### 6.2.1**. Introduction**

* The role of this function is to process each record and check its validity with the following set of identifier mandatory fields discussed earlier
* The rest of fields which are not matching the required format but are not mandatory are taken for processing but are assigned “00” value for to values to be in consistent format

### 6.2.2**.Input**

Record to be processed

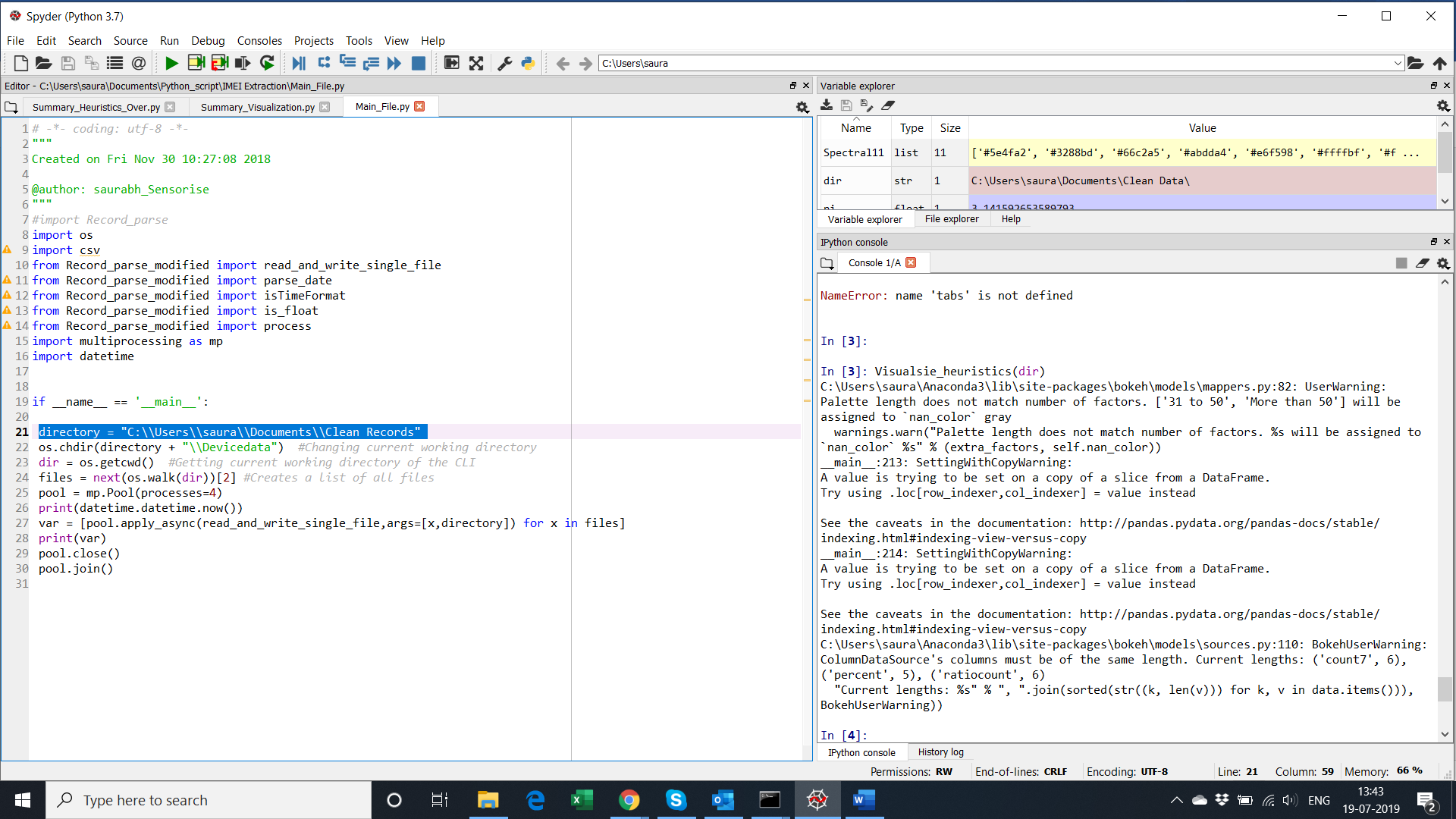
### 6.2.3**. Output**

Following string value is generated after processing of the record:

* True / False: True or false depending upon validity of record
* Good Record/ Error Record: Good record if the validity of mentioned fields with proper consistent values is successful or   
  error record
* Newrecord: it contains all the values which are to be written in track or error file
* IMEI: IMEI number whose record is to be processed
* Moderecord : All the records in form of string whose values or fields are to be written in IMEI file with proper format

7. Procedure to Execute

* The port files whose data is to be processed are multi processed with four core processors with main file specifying function name
* Files to be processed are kept in Device data folder inside clean records sub folder
* The directory has to be mentioned where files are to be stored in **directory** variable i.e in this case directory = "C:\\Users\\saura\\Documents\\Clean Records" in th main file of clean records



* Create One main Folder and 4 subfolders where Track file, Error Summary file and Data files could be stored i.e in this case folders are named as following(see screenshot for ref.)

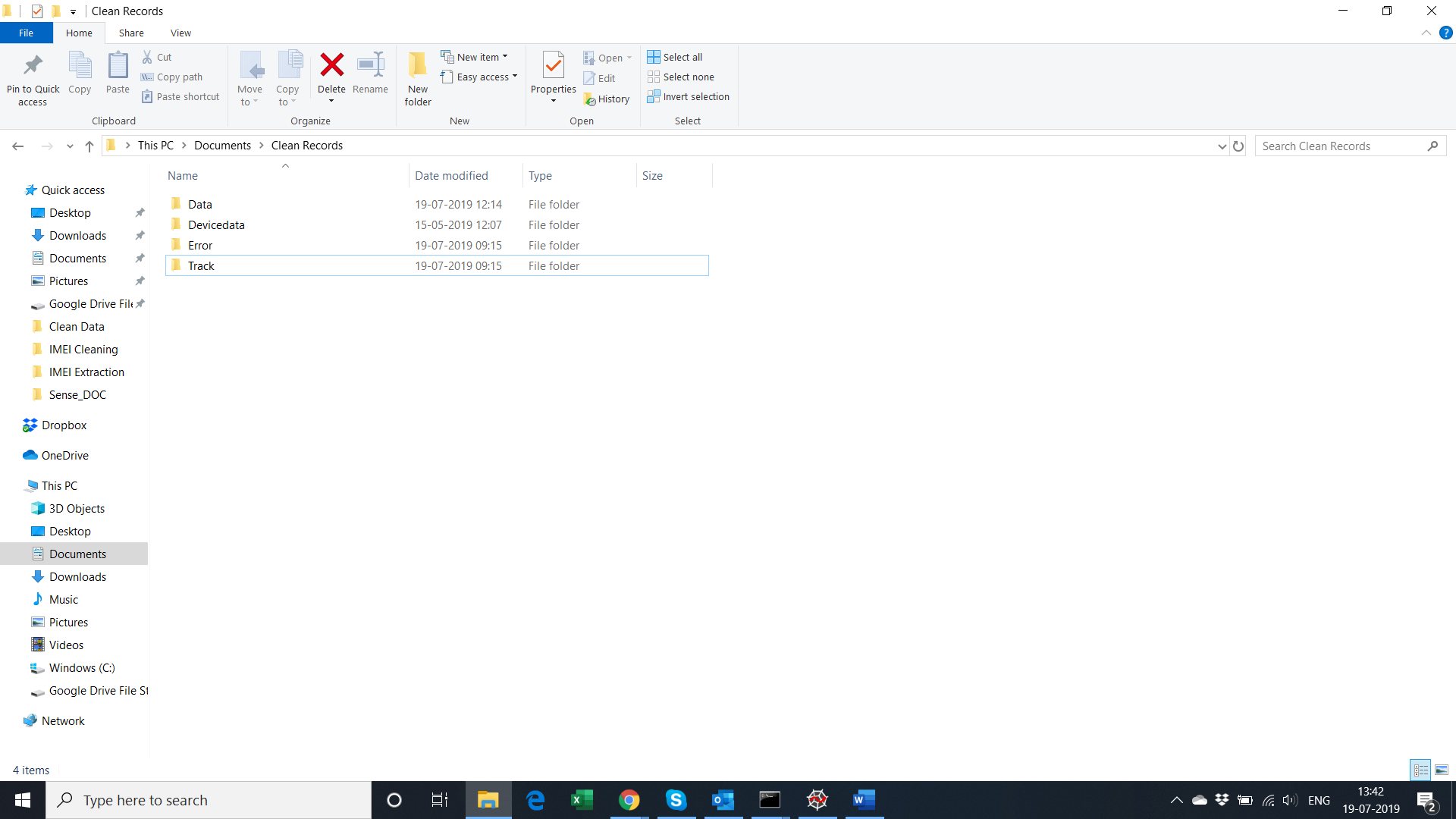




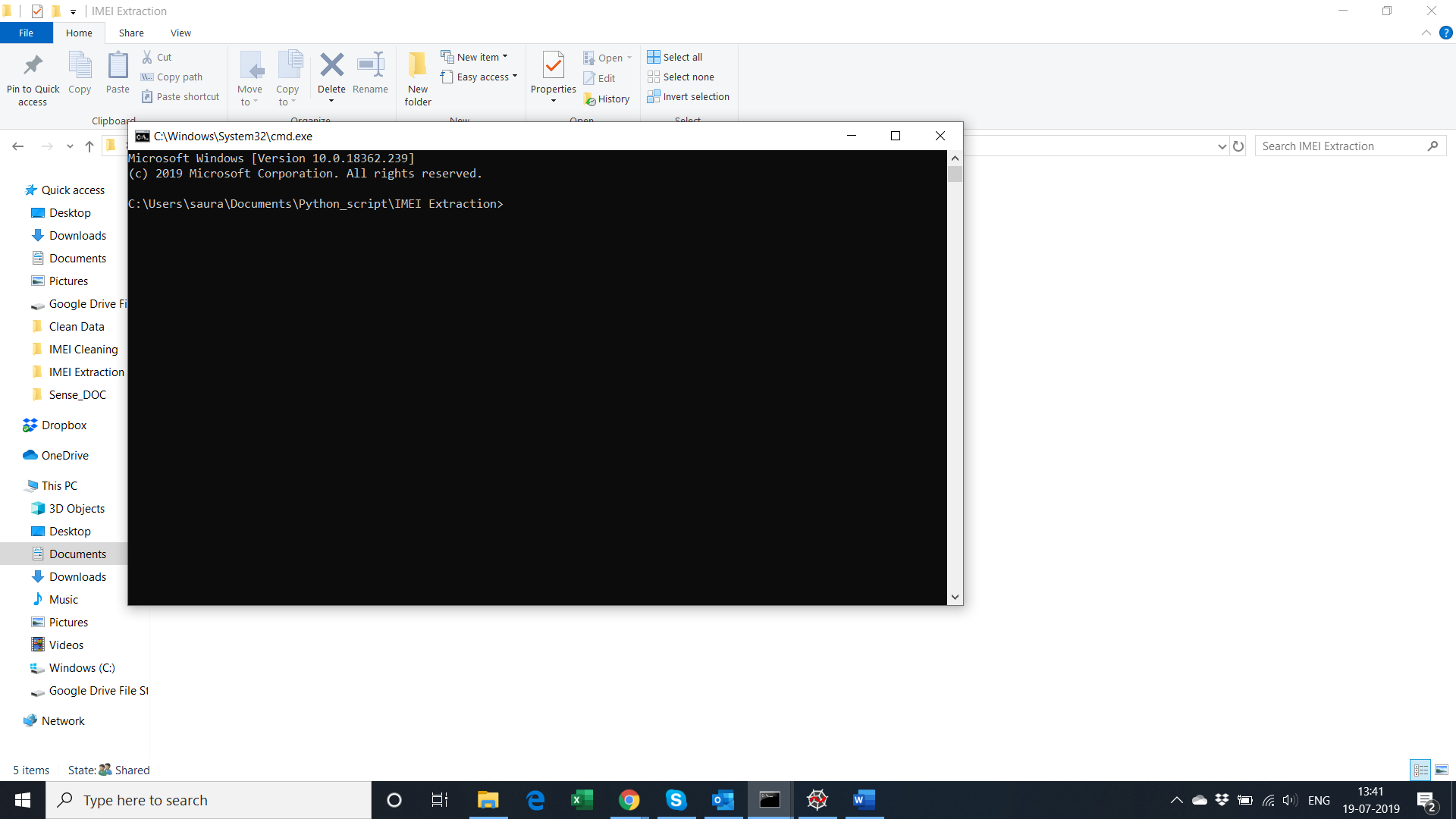








* Open the cmd command line where we have stored the python script of IMEI Extraction and type in cmd – python Main\_file.py(name of file)



8. Files Information

8.1. Input Raw files

|  |  |  |  |
| --- | --- | --- | --- |
| S.no | Port File | Total Files | Size(GB) |
| 1 | 8099 | 31 | 1.2 |
| 2 | 8094 | 31 | 6.2 |
| 3 | 8090 | 31 | 9.10 |
| 4 | 8092 | 31 | 34.1 |
| 5 | 8091 | 31 | 10.6 |

8.2. Output Files

|  |  |  |  |
| --- | --- | --- | --- |
| S.no | File Type | Total Files | Description |
| 1 | Error Summary | 31 \* 5 = 155 | Stored in directory//Error  Contains Error Records of port files |
| 2 | Track | 31 \* 5 = 155 | Stored in directory//Track  Contains Good Records of port files |
| 3 | File Summary | 31 \* 5 = 155 | Stored in directory itself |
| 4 | IMEI | ~15723(approx.) | IMEI records stored in directory//data |