DAEN 690 Capstone Project

Team LEGO

09/19/2021

With the provided UAS Incidents dataset, this document will step by step clean the data and separate the data into standard and non-standard datasets, as well as extract the distance, bearing and location information of the aircraft where possible.

Read the data and look at quick summary stats.

```
file <- "Incidents_Original_Adjusted.csv"
df <- read.csv(file,header = TRUE, stringsAsFactors = FALSE, na.strings=c("NA","N/A","","na","Na","n/a"
setDT(df)
df
```

```
##                    DATE  CALLSIGN   POD PRIMARYCODE SECONDARYCODES
##    1: 2018-01-01T00:58Z    JBU351   DEN         UAS           <NA>
##    2: 2018-01-01T10:46Z   JBU1841   DEN         UAS           <NA>
##    3: 2018-01-01T14:37Z     STAR8   DEN         UAS           <NA>
##    4: 2018-01-01T14:39Z LIFEFLT37   DEN         UAS           <NA>
##    5: 2018-01-01T20:23Z   AWI4292   DEN         UAS           <NA>
##   ---
## 9210: 2021-08-25T17:02Z    UAL967   DEN         UAS           <NA>
## 9211: 2021-08-25T19:25Z   JBU2016   DEN         UAS           <NA>
## 9212: 2021-08-26T00:47Z      <NA>   DEN         UAS           <NA>
## 9213: 2021-08-26T00:50Z      <NA> JATOC         UAS           <NA>
## 9214: 2021-08-26T04:00Z     ALFT3   DEN         UAS           <NA>
##       REPORTINGFACILITY
##    1:               ZNY
##    2:               CLE
##    3:               SLC
##    4:               SLC
##    5:               PHL
##   ---
## 9210:               EWR
## 9211:               BOS
## 9212:               BOS
## 9213:               BOS
## 9214:               BFI
##
##    1:
##    2:
##    3:
##    4:
##    5:
##   ---
## 9210:
## 9211:
## 9212:
## 9213: 2110 EDT / 0110 UTC  8/26/2021\nMASS State PD personnel observed a UAS on the RWY15R final. Ai
## 9214:
##                             ACTYPE
```

```
##    1:            A320, AIRBUS, A-320
##    2:             E190, EMBRAER, 190
##    3:               HELO, HELO, HELO
##    4:               HELO, HELO, HELO
##    5: CRJ2, CANADAIR, Challenger 800
##   ---
## 9210:                          B752
## 9211:                          E190
## 9212:                          <NA>
## 9213:                          <NA>
## 9214:                          EC35
##                                                                      ORIGIN
##    1:                                    MMUN, , CANCUN INTL, MEXICO, CANCUN
##    2: KCLE, CLE, CLEVELAND-HOPKINS INTL, UNITED STATES OF AMERICA, CLEVELAND
##    3:                                                        , VFR, VFR, ,
##    4:                                                        , VFR, VFR, ,
##    5:   KPHL, PHL, PHILADELPHIA INTL, UNITED STATES OF AMERICA, PHILADELPHIA
##   ---
## 9210:               BIKF, KEFLAVIK INTERNATIONAL AIRPORT, REYKJAVIK, ICELAND
## 9211:                KBUF, BUF, BUFFALO NIAGARA INTL, BUFFALO, UNITED STATES
## 9212:                                                                 <NA>
## 9213:                                                                 <NA>
## 9214:                                                 VFR, VFR, VFR, VFR
##                                                                       DEST
##    1:                                                                 <NA>
##    2:                                                                 <NA>
##    3:                                                                 <NA>
##    4:                                                                 <NA>
##    5:                                                                 <NA>
##   ---
## 9210:                KEWR, EWR, NEWARK LIBERTY INTL, NEWARK, UNITED STATES
## 9211: KBOS, BOS, GENERAL EDWARD LAWRENCE LOGAN INTL, BOSTON, UNITED STATES
## 9212:                                                                 <NA>
## 9213:                                                                 <NA>
## 9214:     KBFI, BFI, BOEING FIELD/KING COUNTY INTL, SEATTLE, UNITED STATES
##                                                                    DESTNEW
##    1:              KJFK, JFK, JOHN F KENNEDY INTL, UNITED STATES OF AMERICA, NEW YORK
##    2: KBOS, BOS, GENERAL EDWARD LAWRENCE LOGAN INTL, UNITED STATES OF AMERICA, BOSTON
##    3:                                                         , VFR, VFR, ,
##    4:                                                         , VFR, VFR, ,
##    5:           KMKE, MKE, GENERAL MITCHELL INTL, UNITED STATES OF AMERICA, MILWAUKEE
##   ---
## 9210:                                                                   <NA>
## 9211:                                                                   <NA>
## 9212:                                                                   <NA>
## 9213:                                                                   <NA>
## 9214:                                                                   <NA>
##       IMPACTEDFACILITY OPLVL CEDAR.REMARKS
##    1:               NA    NA          <NA>
##    2:               NA    NA          <NA>
##    3:               NA    NA          <NA>
##    4:               NA    NA          <NA>
##    5:               NA    NA          <NA>
##   ---
```

```
## 9210:               NA    NA          <NA>
## 9211:               NA    NA          <NA>
## 9212:               NA    NA          <NA>
## 9213:               NA    NA          <NA>
## 9214:               NA    NA          <NA>
```

```
# Dimension of data set
dim(df)
```

```
## [1] 9214   14
```

```
# Names of fields
names(df)
```

```
##  [1] "DATE"              "CALLSIGN"          "POD"
##  [4] "PRIMARYCODE"       "SECONDARYCODES"    "REPORTINGFACILITY"
##  [7] "REMARKS"           "ACTYPE"            "ORIGIN"
## [10] "DEST"              "DESTNEW"           "IMPACTEDFACILITY"
## [13] "OPLVL"             "CEDAR.REMARKS"
```

```
# Structure of fields
str(df)
```

```
## Classes 'data.table' and 'data.frame':   9214 obs. of  14 variables:
##  $ DATE             : chr  "2018-01-01T00:58Z" "2018-01-01T10:46Z" "2018-01-01T14:37Z" "2018-01-01T14
##  $ CALLSIGN         : chr  "JBU351" "JBU1841" "STAR8" "LIFEFLT37" ...
##  $ POD              : chr  "DEN" "DEN" "DEN" "DEN" ...
##  $ PRIMARYCODE      : chr  "UAS" "UAS" "UAS" "UAS" ...
##  $ SECONDARYCODES   : chr  NA NA NA NA ...
##  $ REPORTINGFACILITY: chr  "ZNY" "CLE" "SLC" "SLC" ...
##  $ REMARKS          : chr  "Aircraft reported a UAS off the left side, 3 NM NE of CRI VOR while soutl
##  $ ACTYPE           : chr  "A320, AIRBUS, A-320" "E190, EMBRAER, 190" "HELO, HELO, HELO" "HELO, HELO
##  $ ORIGIN           : chr  "MMUN, , CANCUN INTL, MEXICO, CANCUN" "KCLE, CLE, CLEVELAND-HOPKINS INTL,
##  $ DEST             : chr  NA NA NA NA ...
##  $ DESTNEW          : chr  "KJFK, JFK, JOHN F KENNEDY INTL, UNITED STATES OF AMERICA, NEW YORK" "KBOS
##  $ IMPACTEDFACILITY : logi  NA NA NA NA NA NA ...
##  $ OPLVL            : logi  NA NA NA NA NA NA ...
##  $ CEDAR.REMARKS    : chr  NA NA NA NA ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

```
# Summary of fields
summary(df)
```

```
##      DATE             CALLSIGN             POD            PRIMARYCODE
##  Length:9214        Length:9214        Length:9214        Length:9214
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##  SECONDARYCODES     REPORTINGFACILITY    REMARKS              ACTYPE
##  Length:9214        Length:9214        Length:9214        Length:9214
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##     ORIGIN              DEST              DESTNEW          IMPACTEDFACILITY
##  Length:9214        Length:9214        Length:9214        Mode:logical
##  Class :character   Class :character   Class :character   NA's:9214
##  Mode  :character   Mode  :character   Mode  :character
##   OPLVL         CEDAR.REMARKS
##  Mode:logical   Length:9214
```

```
##  NA's:9214      Class :character
##                 Mode  :character
```

Split up CEDAR Remarks

```
df_2 <- separate(data=df, col=CEDAR.REMARKS, into=c("EVENTTYPE.CEDAR","STATUS.CEDAR","MORID.CEDAR","FAC

df_2$EVENTTYPE.CEDAR <- gsub("CEDAR - Event Type: ","",df_2$EVENTTYPE.CEDAR)
df_2$STATUS.CEDAR <- gsub("Status: ","",df_2$STATUS.CEDAR)
df_2$MORID.CEDAR <- gsub("MOR ID: ","",df_2$MORID.CEDAR)
df_2$FACILITY.CEDAR <- gsub("Facility: ","",df_2$FACILITY.CEDAR)
df_2$EVENTDATE.CEDAR <- gsub("Event Date: ","",df_2$EVENTDATE.CEDAR)
df_2$UTCTIME.CEDAR <- gsub("UTC Time: ","",df_2$UTCTIME.CEDAR)
df_2$UTCTIME24.CEDAR <- gsub("UTC Time 24 HR Format: ","",df_2$UTCTIME24.CEDAR)
df_2$CALENDARDATE.CEDAR <- gsub("Calendar Date: ","",df_2$CALENDARDATE.CEDAR)
df_2$NEARESTAIRPORT.CEDAR <- gsub("Nearest Airport: ","",df_2$NEARESTAIRPORT.CEDAR)
df_2$METAR.CEDAR <- gsub("METAR: ","",df_2$METAR.CEDAR)
df_2$POTENTIALLYSIGNIFICANT.CEDAR <- gsub("Potentially Significant: ","",df_2$POTENTIALLYSIGNIFICANT.CE
df_2$CALLSIGN.CEDAR <- gsub("Callsign: ","",df_2$CALLSIGN.CEDAR)
df_2$ACTYPE.CEDAR <- gsub("A/C Type: ","",df_2$ACTYPE.CEDAR)
df_2$IFRVFR.CEDAR <- gsub("IFR / VFR: ","",df_2$IFRVFR.CEDAR)
df_2$AUTHCERT.CEDAR <- gsub("Certificate of Authorization: ","",df_2$AUTHCERT.CEDAR)
df_2$AIRSPACECLASS.CEDAR <- gsub("Airspace Class: ","",df_2$AIRSPACECLASS.CEDAR)
df_2$ACLOCATION.CEDAR <- gsub("A/C Location F/R/D: ","",df_2$ACLOCATION.CEDAR)
df_2$ACALTITUDE.CEDAR <- gsub("A/C Altitude: ","",df_2$ACALTITUDE.CEDAR)
df_2$ACHEADING.CEDAR <- gsub("A/C Heading: ","",df_2$ACHEADING.CEDAR)
df_2$RELATIVECLOCKPOSITION.CEDAR <- gsub("Relative Clock Position: ","",df_2$RELATIVECLOCKPOSITION.CEDA
df_2$UASREGISTRATIONNUM.CEDAR <- gsub("UAS Registration #: ","",df_2$UASREGISTRATIONNUM.CEDAR)
df_2$UASLONG.CEDAR <- gsub("UAS Longitude: ","",df_2$UASLONG.CEDAR)
df_2$UASLAT.CEDAR <- gsub("UAS Latitude: ","",df_2$UASLAT.CEDAR)
df_2$UASTYPE.CEDAR <- gsub("UAS Type: ","",df_2$UASTYPE.CEDAR)
df_2$UASFORMATION.CEDAR <- gsub("UAS Formation: ","",df_2$UASFORMATION.CEDAR)
df_2$CLOSESTPROXIMITY.CEDAR <- gsub("Closest Proximity \\(feet\\): ","",df_2$CLOSESTPROXIMITY.CEDAR)
df_2$UASWEIGHTGT55.CEDAR <- gsub("UAS Weight Exceeds 55lbs: ","",df_2$UASWEIGHTGT55.CEDAR)
df_2$UASDIM.CEDAR <- gsub("UAS Dimensions \\(feet\\): ","",df_2$UASDIM.CEDAR)
df_2$UASFWROTOR.CEDAR <- gsub("UAS Fixed Wing/Rotorcraft: ","",df_2$UASFWROTOR.CEDAR)
df_2$UASACTIVITYRISK.CEDAR <- gsub("UAS Activity Risk: ","",df_2$UASACTIVITYRISK.CEDAR)
df_2$UASCOLOR.CEDAR <- gsub("UAS Color: ","",df_2$UASCOLOR.CEDAR)
df_2$PILOTREPORTEDNMAC.CEDAR <- gsub("Pilot Reported as NMAC: ","",df_2$PILOTREPORTEDNMAC.CEDAR)
df_2$TCASRA.CEDAR <- gsub("TCAS RA: ","",df_2$TCASRA.CEDAR)
df_2$LEOCONTACT.CEDAR <- gsub("Law Enforcement Contact Info: ","",df_2$LEOCONTACT.CEDAR)
df_2$SUMMARY.CEDAR <- gsub("Summary: ","",df_2$SUMMARY.CEDAR)
df_2$QAFINDINGS.CEDAR <- gsub("QA Findings: ","",df_2$QAFINDINGS.CEDAR)
```

General Cleaning

```
df_2$REMARKS <- gsub("ACFT","Aircraft",df_2$REMARKS)
df_2$REMARKS <- gsub("(M|m)iles?","NM",df_2$REMARKS)
df_2$SUMMARY.CEDAR <- gsub("(M|m)iles?","NM",df_2$SUMMARY.CEDAR)
df_2$REMARKS <- gsub("of the","of",df_2$REMARKS)
df_2$REMARKS <- gsub("(Runway|runway|RUNWAY)","RWY",df_2$REMARKS)
df_2$REMARKS <- gsub("RY","RWY",df_2$REMARKS)
df_2$REMARKS <- gsub("UAS","uas",df_2$REMARKS)
```

```
df_2$REMARKS <- gsub("(NM)([A-Z]{1,3})","\\1 \\2",df_2$REMARKS)
df_2$REMARKS <- gsub("([0-9]*\\-*\\/*[0-9]*\\.*[0-9]\\.*[0-9]*)(NM)","\\1 \\2",df_2$REMARKS)
df_2$REMARKS <- gsub("(of)([A-Z]{3,4})","\\1 \\2",df_2$REMARKS)
df_2$REMARKS <- gsub("([A-Z]{1,3})(of)","\\1 \\2",df_2$REMARKS)
df_2$REMARKS <- gsub("([A-Z]{1,3})(\\s[A-Z]{3,4}$)","\\1 of\\2",df_2$REMARKS)
df_2$REMARKS <- gsub("(RWY)([0-9]{1,2}[L|R|C]?)","\\1 \\2",df_2$REMARKS)
df_2$REMARKS <- gsub("(RWY\\s)(\\d(?!\\d))","\\10\\2",df_2$REMARKS,perl=T)
df_2$REMARKS <- gsub("(South|south|SOUTH)","S",df_2$REMARKS)
df_2$REMARKS <- gsub("(East|east|EAST)","E",df_2$REMARKS)
df_2$REMARKS <- gsub("(North|north|NORTH)","N",df_2$REMARKS)
df_2$REMARKS <- gsub("(West|west|WEST)","W",df_2$REMARKS)
df_2$REMARKS <- gsub("  "," ",df_2$REMARKS)
```

Adjust date column to be readable as date/time

```
dt <- df_2$DATE
dtparts <- t(as.data.frame(strsplit(dt,"T")))
dtparts[,2] <- substr(dtparts[,2],1,5)
dateonly <- dtparts[,1]
timeonly <- dtparts[,2]

df_2 <- cbind(timeonly,df_2)
df_2 <- cbind(dateonly,df_2)
df_2$dateonly <- as.Date(df_2$dateonly)
```

```
# removal of columns that are completely empty
df_2 <- df_2[,c("IMPACTEDFACILITY","OPLVL","UASTYPE.CEDAR","UASACTIVITYRISK.CEDAR","LEOCONTACT.CEDAR"):=

# removal of rows with codes unnecessary to project; add to exception file
pattern <- c("ADMIN","AIRCRAFT ACCIDENT","AIRPORT","ATC FACILITY","C-UAS","DISTURB","EQUIPMENT","HORNET
df_exp <- df_2[grep(paste(pattern,collapse="|"),df_2$PRIMARYCODE)]
df_2 <- df_2[!grep(paste(pattern,collapse="|"),df_2$PRIMARYCODE)]
df_exp <- rbind(df_exp,df_2[grep(paste(pattern,collapse="|"),df_2$SECONDARYCODES)])
df_2 <- anti_join(df_2,df_exp)
```

```
## Joining, by = c("dateonly", "timeonly", "DATE", "CALLSIGN", "POD", "PRIMARYCODE", "SECONDARYCODES",
```

```
df_exp$DATASET <- "EXCEPTION - UNRELATED CODES"
```

Find standard format remarks

```
df_stob <- df_2[grepl("Aircraft observed a",df_2$REMARKS)]
df_stre <- df_2[grepl("Aircraft reported a",df_2$REMARKS)]

df_3 <- rbind(df_stob,df_stre)

# group the non-standard format ones together
df_nstob <- df_2[!grepl("Aircraft observed a",df_2$REMARKS)]
df_nstre <- df_nstob[!grepl("Aircraft reported a",df_nstob$REMARKS)]
df_4 <- df_nstre
```

Write to CSV

```
write.csv(df_2,"C:\\Users\\maygo\\OneDrive\\Documents\\DAEN690-FAA-UAS\\00-Incidents_Cleaned.csv",row.na
write.csv(df_3,"C:\\Users\\maygo\\OneDrive\\Documents\\DAEN690-FAA-UAS\\01-Incidents_Cleaned_Standard.cs
write.csv(df_4,"C:\\Users\\maygo\\OneDrive\\Documents\\DAEN690-FAA-UAS\\02-Incidents_Cleaned_NonStandar
```

Extract location

```
# uas lat/long can be extracted from cedar:
df_2$DATASET <- ifelse(df_2$UASLAT.CEDAR != "NA" & df_2$UASLONG.CEDAR != "NA","CEDAR LAT/LONG",NA)
df_final <- df_2[grep("CEDAR LAT/LONG",df_2$DATASET)]
df_2 <- df_2[!grep("CEDAR LAT/LONG",df_2$DATASET)]

# remove FRZ
df_frz <- df_2[grep("FRZ",df_2$REMARKS)]
df_frz$DATASET <- "EXCEPTION - FRZ"
df_exp <- rbind(df_exp,df_frz)
df_2 <- df_2[!grep("FRZ",df_2$REMARKS)]
remove(df_frz)

#extract all designated points
df_2$UASLOCATION <- str_extract(df_2$REMARKS,'\\b\\d[.|-|/]*\\d*\\s?(nm|NM)\\s?(N|S|E|W|NW|NE|SW|SE|SSE
df_dp <- df_2[!is.na(df_2$UASLOCATION)]
df_dp$DATASET <- "DESIGNATED POINT"
df_dp_dups <- df_dp[duplicated(df_dp[,c("dateonly","UASLOCATION")]),]
df_dp <- distinct(df_dp,dateonly,UASLOCATION,.keep_all=TRUE)
df_final <- rbind(df_final,df_dp,fill=TRUE)
df_2 <- df_2[is.na(df_2$UASLOCATION)]
df_exp <- rbind(df_exp,df_dp_dups,fill=TRUE)
remove(df_dp,df_dp_dups)

#extract all navaid
pattern <- c("VOR","vor","NDB","ndb")
df_3 <- df_2[grep(paste(pattern,collapse="|"),df_2$REMARKS)]
df_3$DATASET <- "NAVAID"
df_3$UASLOCATION <- str_extract(df_3$REMARKS,'\\b\\d[.|-|/]*\\d*\\s?(nm|NM)\\s?(N|S|E|W|NW|NE|SW|SE|SSE
df_navaid <- df_3[!is.na(df_3$UASLOCATION)]
df_navaid_dups <- df_navaid[duplicated(df_navaid[,c("dateonly","UASLOCATION")]),]
df_navaid <- distinct(df_navaid,dateonly,UASLOCATION,.keep_all=TRUE)
df_navaid_na <- df_3[is.na(df_3$UASLOCATION)]
df_navaid_na$UASLOCATION <- df_navaid_na$REPORTINGFACILITY
df_navaid_dups_na <- df_navaid_na[duplicated(df_navaid_na[,c("dateonly","UASLOCATION")]),]
df_navaid_na <- distinct(df_navaid_na,dateonly,UASLOCATION,.keep_all=TRUE)

df_navaid_over <- df_navaid_na[grepl("over ",df_navaid_na$REMARKS)]
df_navaid_na <- df_navaid_na[!grepl("over ",df_navaid_na$REMARKS)]
df_navaid_over$UASLOCATION <- str_extract(df_navaid_over$REMARKS,'(?<=over\\s)([A-Z]{3})')
df_navaid_over$DATASET <- "NAVAID DIRECTLY OVER"
df_exp <- rbind(df_exp,df_navaid_na,df_navaid_dups,df_navaid_dups_na)
df_final <- rbind(df_final,df_navaid,df_navaid_over)
df_2 <- df_2[!grep(paste(pattern,collapse="|"),df_2$REMARKS)]
remove(df_navaid,df_navaid_dups,df_navaid_dups_na,df_navaid_na,df_navaid_over,df_3)
```

```
#extract all runway
df_2 <- df_2[!grep(paste(pattern,collapse="|"),df_2$REMARKS)]
df_2$RWYLOCATION <- str_extract(df_2$REMARKS,'RWY\\s?[0-9]{1,2}[L|R|C]?')
df_rwy <- df_2[!is.na(df_2$RWYLOCATION)]
df_rwy$DATASET <- "RUNWAY"
df_rwy$UASLOCATION <- str_extract(df_rwy$REMARKS,'\\b\\d[.|-|/]*\\d*\\s?(nm|NM)\\s*(N|S|E|W|NW|NE|SW|SE
df_22 <- df_rwy[is.na(df_rwy$UASLOCATION)]
df_22$UASLOCATION <- str_extract(df_22$REMARKS,'\\b\\d[.|-|/]*\\d*\\s?(nm|NM)\\s*(N|S|E|W|NW|NE|SW|SE|S
df_rwy2 <- df_22[!is.na(df_22$UASLOCATION)]
df_rwy2$UASLOCATION <- paste(df_rwy2$UASLOCATION,df_rwy2$REPORTINGFACILITY)
df_rwy_na <- df_22[is.na(df_22$UASLOCATION)]
df_rwy_na$UASLOCATION <- df_rwy_na$REPORTINGFACILITY
df_rwy <- df_rwy[!is.na(df_rwy$UASLOCATION)]
df_rwy <- rbind(df_rwy,df_rwy2,fill=TRUE)
df_rwy_dups <- df_rwy[duplicated(df_rwy[,c("dateonly","UASLOCATION")]),]
df_rwy <- distinct(df_rwy,dateonly,UASLOCATION,.keep_all=TRUE)
df_exp <- rbind(df_exp,df_rwy_na,df_rwy_dups,fill=TRUE)
df_final <- rbind(df_final,df_rwy,fill=TRUE)
df_2 <- df_2[is.na(df_2$RWYLOCATION)]
remove(df_rwy,df_22,df_rwy2,df_rwy_na,df_rwy_dups)

#extract all airports
df_2$UASLOCATION <- str_extract(df_2$REMARKS,'\\b\\d[.|-|/]*\\d*\\s?(nm|NM)\\s?(N|S|E|W|NW|NE|SW|SE|SSE
df_airport <- df_2[!is.na(df_2$UASLOCATION)]
df_2 <- df_2[is.na(df_2$UASLOCATION)]
df_2$UASLOCATION <- str_extract(df_2$REMARKS,'\\b\\d[.|-|/]*\\d*\\s?(N|S|E|W|NW|NE|SW|SE|SSE|SSW|SNE|SN
df_airport2 <- df_2[!is.na(df_2$UASLOCATION)]
df_airport2$UASLOCATION <- gsub("(\\d[.|-|/]*\\d*\\s)(N|S|E|W|NW|NE|SW|SE|SSE|SSW|SNE|SNW|NNE|NSE|NNE|N
df_airport <- rbind(df_airport,df_airport2)
df_airport$DATASET <- "AIRPORT"
df_airport_dups <- df_airport[duplicated(df_airport[,c("dateonly","UASLOCATION")]),]
df_airport <- distinct(df_airport,dateonly,UASLOCATION,.keep_all=TRUE)

df_airport_over <- df_2[is.na(df_2$UASLOCATION)]
df_airport_over$UASLOCATION <- str_extract(df_airport_over$REMARKS,'(?<=over\\s)([A-Z]{3})')
df_2 <- df_airport_over[is.na(df_airport_over$UASLOCATION)]
df_airport_over <- df_airport_over[!is.na(df_airport_over$UASLOCATION)]
df_airport_over$DATASET <- "AIRPORT DIRECTLY OVER"

df_exp <- rbind(df_exp,df_airport_dups)
df_final <- rbind(df_final,df_airport,df_airport_over)
remove(df_airport,df_airport_dups,df_airport_over,df_airport2)

# fixed radial distance
df_2$UASLOCATION <- str_extract(df_2$REMARKS,'[A-Z]{3}\\d{6}')
df_frd <- df_2[!is.na(df_2$UASLOCATION)]
df_frd$location <- str_extract(df_frd$UASLOCATION,'^[A-Z]{3}')
df_frd$heading <- str_extract(df_frd$UASLOCATION,'[0-9]{3}([0-9]{3})$')
df_frd$heading <- str_extract(df_frd$heading,'^[0-9]{3}')
df_frd$nm <- str_extract(df_frd$UASLOCATION,'[0-9]{3}$')
df_frd$UASLOCATION <- paste(df_frd$nm,"NM",df_frd$heading,"of",df_frd$location)
df_frd$DATASET <- "FIX RADIAL"
df_frd <- df_frd[,c("location","heading","nm"):=NULL]
```

```
df_frd_dups <- df_frd[duplicated(df_frd[,c("dateonly","UASLOCATION")]),]
df_frd <- distinct(df_frd,dateonly,UASLOCATION,.keep_all=TRUE)
df_2 <- df_2[is.na(df_2$UASLOCATION)]
df_exp <- rbind(df_exp,df_frd_dups)
df_final <- rbind(df_final,df_frd)
remove(df_frd,df_frd_dups)

#extract all airports - CEDAR
df_2$UASLOCATION <- str_extract(df_2$SUMMARY.CEDAR,'\\b\\d[.|-|/]*\\d*\\s?(nm|NM)\\s?(N|S|E|W|NW|NE|SW|S
df_cedarair <- df_2[!is.na(df_2$UASLOCATION)]
df_2 <- df_2[is.na(df_2$UASLOCATION)]
df_2$UASLOCATION <- str_extract(df_2$SUMMARY.CEDAR,'\\b\\d[.|-|/]*\\d*\\s?(N|S|E|W|NW|NE|SW|SE|SSE|SSW|S
df_cedarair2 <- df_2[!is.na(df_2$UASLOCATION)]
df_cedarair2$UASLOCATION <- gsub("(\\d[.|-|/]*\\d*\\s?)(N|S|E|W|NW|NE|SW|SE|SSE|SSW|SNE|SNW|NNE|NSE|NNE
df_cedarair <- rbind(df_cedarair,df_cedarair2)
df_cedarair$DATASET <- "AIRPORT - CEDAR"
df_cedarair_dups <- df_cedarair[duplicated(df_cedarair[,c("dateonly","UASLOCATION")]),]
df_cedarair <- distinct(df_cedarair,dateonly,UASLOCATION,.keep_all=TRUE)
df_2 <- df_2[is.na(df_2$UASLOCATION)]
df_exp <- rbind(df_exp,df_cedarair_dups)
df_final <- rbind(df_final,df_cedarair)
remove(df_cedarair,df_cedarair2,df_cedarair_dups)

# fixed radial distance - CEDAR
df_2$UASLOCATION <- str_extract(df_2$SUMMARY.CEDAR,'[A-Z]{3}\\d{6}')
df_cedarfrd <- df_2[!is.na(df_2$UASLOCATION)]
df_cedarfrd$location <- str_extract(df_cedarfrd$UASLOCATION,'^[A-Z]{3}')
df_cedarfrd$heading <- str_extract(df_cedarfrd$UASLOCATION,'[0-9]{3}([0-9]{3})$')
df_cedarfrd$heading <- str_extract(df_cedarfrd$heading,'^[0-9]{3}')
df_cedarfrd$nm <- str_extract(df_cedarfrd$UASLOCATION,'[0-9]{3}$')
df_cedarfrd$UASLOCATION <- paste(df_cedarfrd$nm,"NM",df_cedarfrd$heading,"of",df_cedarfrd$location)
df_cedarfrd$DATASET <- "FIX RADIAL - CEDAR"
df_cedarfrd <- df_cedarfrd[,c("location","heading","nm"):=NULL]
df_cedarfrd_dups <- df_cedarfrd[duplicated(df_cedarfrd[,c("dateonly","UASLOCATION")]),]
df_cedarfrd <- distinct(df_cedarfrd,dateonly,UASLOCATION,.keep_all=TRUE)
df_2 <- df_2[is.na(df_2$UASLOCATION)]
df_exp <- rbind(df_exp,df_cedarfrd_dups)
df_final <- rbind(df_final,df_cedarfrd)
remove(df_cedarfrd,df_cedarfrd_dups)


#extract all else
df_2$UASLOCATION <- str_extract(df_2$REMARKS,'\\b\\d[.|-|/]*\\d*\\s?(nm|NM)\\s?(N|S|E|W|NW|NE|SW|SE|SSE
df_unk <- df_2[!is.na(df_2$UASLOCATION)]
df_unk$UASLOCATION <- paste(df_unk$UASLOCATION,df_unk$REPORTINGFACILITY)
df_2 <- df_2[is.na(df_2$UASLOCATION)]
df_2$UASLOCATION <- str_extract(df_2$REMARKS,'\\b\\d[.|-|/]*\\d*\\s?(N|S|E|W|NW|NE|SW|SE|SSE|SSW|SNE|SN
df_repfac <- df_2[!is.na(df_2$UASLOCATION)]
df_repfac$UASLOCATION <- gsub("(\\d[.|-|/]*\\d*\\s?)(N|S|E|W|NW|NE|SW|SE|SSE|SSW|SNE|SNW|NNE|NSE|NNE|NN
df_repfac$UASLOCATION <- paste(df_repfac$UASLOCATION,df_repfac$REPORTINGFACILITY)
df_unk <- rbind(df_unk,df_repfac)
df_unk$DATASET <- "AIRPORT(?)"
df_unk_dups <- df_unk[duplicated(df_unk[,c("dateonly","UASLOCATION")]),]
```

```
df_unk <- distinct(df_unk,dateonly,UASLOCATION,.keep_all=TRUE)
df_unk_na <- df_2[is.na(df_2$UASLOCATION)]
df_unk_na$UASLOCATION <- df_unk_na$REPORTINGFACILITY
df_unk_na$DATASET <- "EXCEPTION"
df_2 <- df_2[is.na(df_2$UASLOCATION)]
df_exp <- rbind(df_exp,df_unk,df_unk_dups,df_unk_na)
remove(df_repfac,df_unk,df_unk_dups,df_unk_na)


df_final <- arrange(df_final,DATE)

write.csv(df_exp,paste("DAEN690_ExceptionFile.csv",sep=""),row.names=FALSE)
write.csv(df_final,paste("DAEN690_CompletedIncidents.csv",sep=""),row.names=FALSE)
```