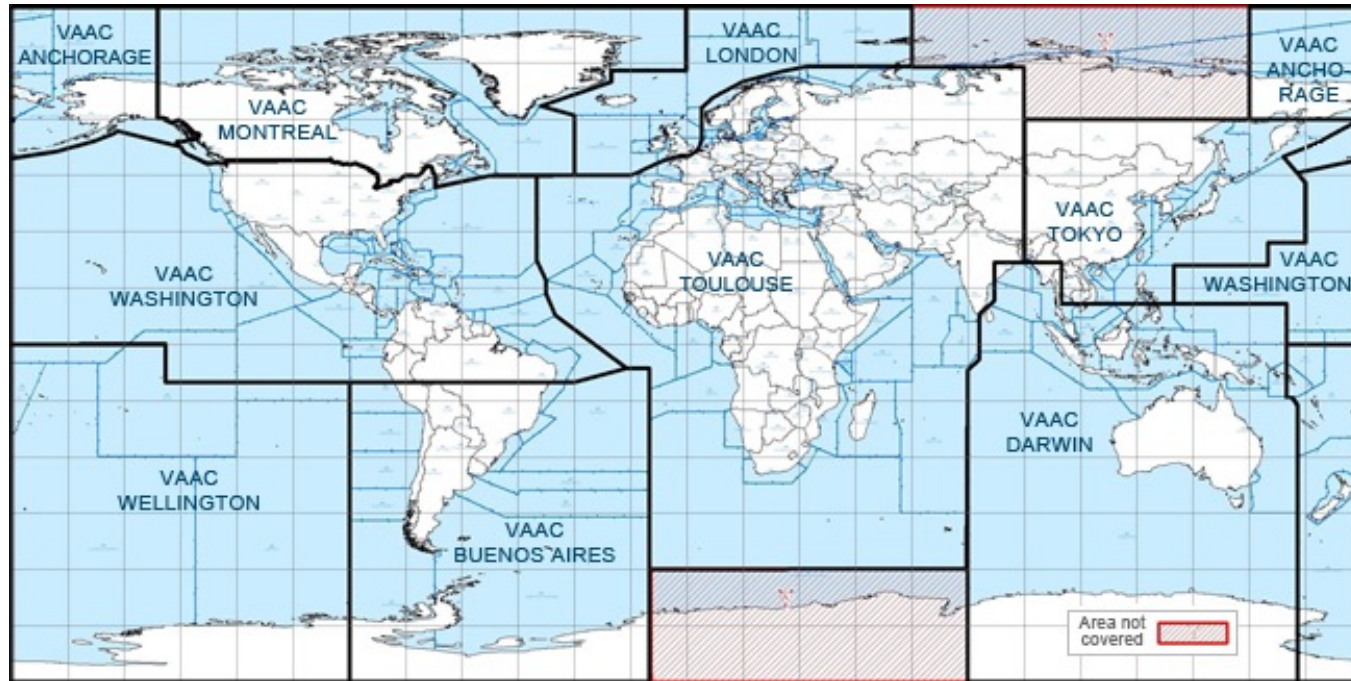


APPLICATIONS USE CASES

Volcanic Ash Advisories

Volcanic Ash Advisories



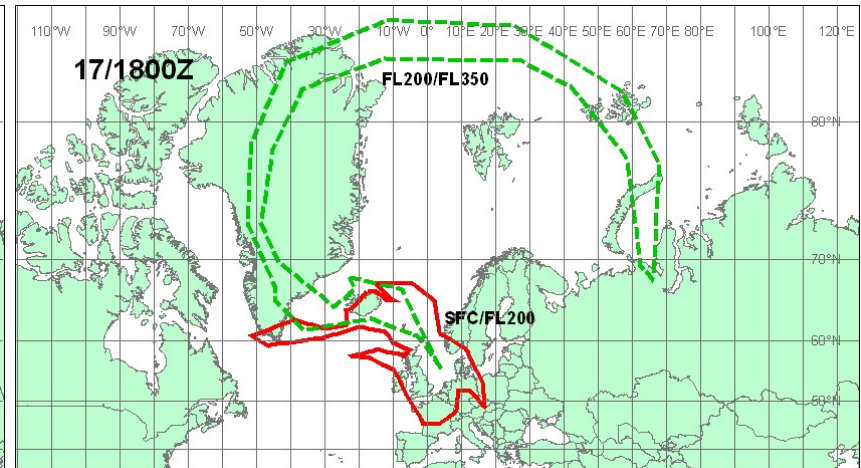
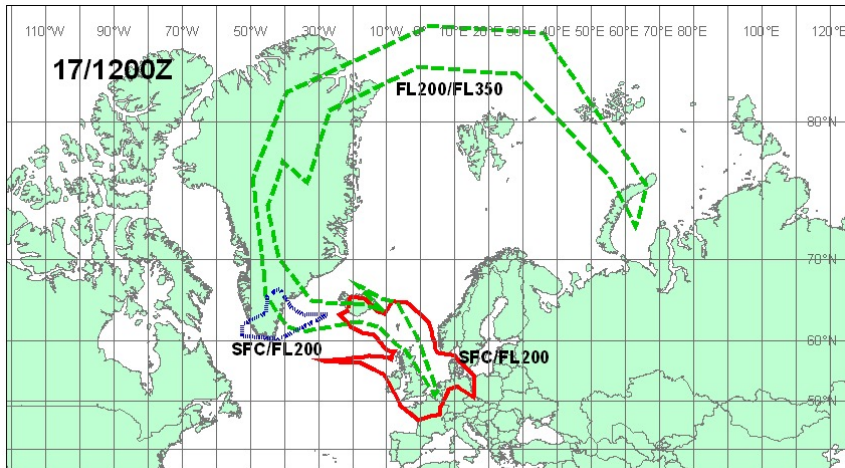
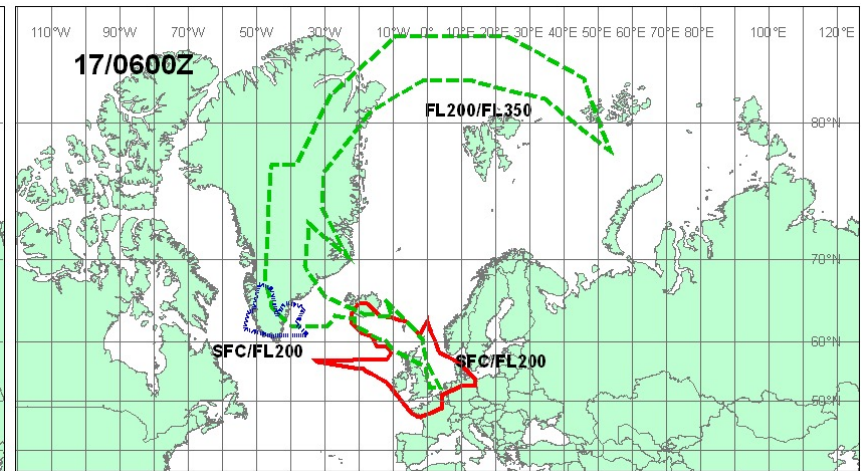
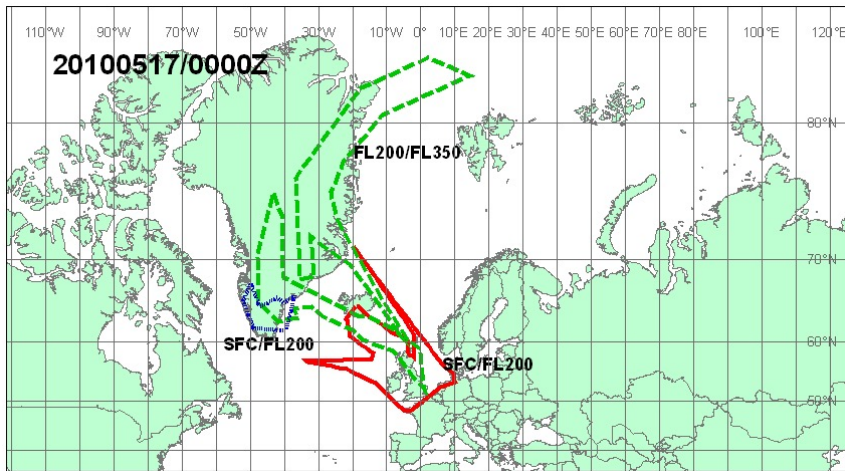
Information Sources:
Pilot reports
Volcano observatories
Satellite imagery

Volcanic Ash Advisory Centres:
Initiate dispersal simulations
Validate simulation results
Provide volcanic ash advisory

Aviation Sector:
Use information to
inform flight paths

Volcanic Ash Advisories

Forecast ash dispersion



VA ADVISORY
DTG: 20100517/0000Z
VAAC: LONDON
VOLCANO:
EYJAFJALLAJOKULL
PSN: N6338 W01937
AREA: ICELAND

SUMMIT ELEV: 1666M
ADVISORY NR: 2010/126
INFO SOURCE: ICELAND MET OFFICE
AVIATION COLOUR CODE: RED
ERUPTION DETAILS: ERUPTION CONTINUING
TO AN ESTIMATED HEIGHT OF FL230 TO FL260,
OCNL FL290.

RMK: A SEPARATE AREA OF SFC TO FL200 OVER GREENLAND IS
MARKED IN BLUE ON GRAPHICS.
NXT ADVISORY: 20100517/0600Z

text providing key information on
the volcano, eruption and source of

Volcanic Ash Advisories

IDD41290

VA ADVISORY

DTG: 20160106/0330Z

VAAC: DARWIN

VOLCANO: SINABUNG 261080

PSN: N0310 E09824

AREA: INDONESIA

SUMMIT ELEV: 2460M

ADVISORY NR: 2016/1

INFO SOURCE: GROUND REPORTS, HIMAWARI-8

AVIATION COLOUR CODE: RED

ERUPTION DETAILS: MINOR VA/STEAM EMISSION TO FL140

EXTENDING TO THE WEST

OBS VA DTG: 06/0330Z

OBS VA CLD: SFC/FL140 N0314 E09824 - N0310 E09810 - N0305

E09811 - N0304 E09817 - N0310 E09826 MOV SW 5KT

FCST VA CLD +6 HR: 06/0930Z SFC/FL140 N0314 E09824 - N0313

E09811 - N0304 E09754 - N0248 E09738 - N0235 E09737 - N0222

E09748 - N0237 E09809 - N0309 E09827

FCST VA CLD +12 HR: 06/1530Z NO VA EXP

FCST VA CLD +18 HR: 06/2130Z NO VA EXP

RMK: VA EXPECTED TO DISSIPATE WITHING 12 HRS.

NXT ADVISORY: NO LATER THAN 20160106/0930Z

Copyright Commonwealth of Australia 2011, Bureau of Meteorology (ABN 92 637 533 532). Users of these web pages are deemed to have read and accepted the conditions described in the Copyright, Disclaimer, and Privacy statements

volcano_name
colour_code
advice_date,
advice_time
time, date, fl,

Volcanic Ash Advisories

Each advisory constitutes a text file containing the following information:

VA ADVISORY

DTG:	20150223/2302Z	[Date and time issuance of advisory]
VAAC:	WELLINGTON	[Responsible VAAC]
VOLCANO:	AMBRYM 257040	[Volcano name and no]
PSN:	S1615 E16807	[Coordinates of volcano]
AREA:	VANUATU	[Volcano area]
SUMMIT ELEV:	1334M	[Elevation of volcano summit]
ADVISORY NR:	2015/17	[Advisory number]
INFO SOURCE:	PIREPS AND SATELLITE SO2 IMAGERY	[Source of information]
AVIATION COLOUR CODE:	UNKNOWN	[Typically provided by local VO's]
ERUPTION DETAILS:	NO OBSERVATIONS OF SIGNIFICANT ASH SINCE PIREP AT 212100Z OF ASH TO 8000FT.	[Information regarding information]
OBS VA DTG:	23/2300Z	[Date and time of observed ash]
EST VA CLD:	VA NOT IDENTIFIABLE FM SATELLITE	[Estimated height (FL) and extent of cloud]
FCST VA CLD+6 HR:	24/0500Z SFC/FL600 NO VA EXP	[Forecast of plume height & extent in 6 hrs]
FCST VA CLD+12 HR:	24/1100Z SFC/FL600 NO VA EXP	[Forecast of plume height & extent in 12 hrs]
FCST VA CLD+18 HR:	24/1700Z SFC/FL600 NO VA EXP	[Forecast of plume height & extent in 18 hrs]
RMK:	SATELLITE IMAGERY SHOWS NO SIGN OF ASH. NO FURTHER ADVISORIES WILL BE ISSUED UNTIL CONFIRMED REPORTS RECEIVED.	[Further info. on eruption and observations]
NXT ADVISORY:	NO FURTHER ADVISORIES=	[Time of next advisory]

Volcanic Ash Advisories

- An “entry”/”row” per report, gathered by “Volcano ID”, with 22 “variables”/”columns”

0502-09	colour_code	fl_eruption_L2	no_id_satelite	fl_eruption_L1	file_name	id	confirmed	info_source		issued_time	is	
	yellow		-1		Darwin/2006/IDD41300.200607130420.txt	0502-09	1	RVO/Sat/Imagery		0420	20	
	yellow		-1		Darwin/2006/IDD41300.200607140356.txt	0502-09	1	RVO/MTSAT-1R/TERRA/MODIS		0356	20	
0502-08	colour_code	fl_eruption_L2	no_id_satelite	fl_eruption_L1	file_name	id	confirmed	info_source	issued_time	issued_date	vaac_obs	r
	orange		0	140	Darwin/2012/IDD41290.201207101643.txt	0502-08	1	MTSAT-2	1643	20120710	Darwin	C
	orange		0	140	Darwin/2012/IDD41290.201207102229.txt	0502-08	1	MTSAT-2	2229	20120710	Darwin	C
	orange		2	140	Darwin/2012/IDD41290.201207110417.txt	0502-08	1	MTSAT-2	0417	20120711	Darwin	C
	red		0	400	Darwin/2012/IDD41295.201205030543.txt	0502-08	1	MTSAT	0542	20120503	Darwin	C
	red		0	450	Darwin/2012/IDD41295.201205030608.txt	0502-08	1	MTSAT-2	0607	20120503	Darwin	C
	orange		0	450	Darwin/2012/IDD41295.201205030722.txt	0502-08	1	MTSAT-2	0721	20120503	Darwin	C
	colour_code	fl_eruption_L2	no_id_satelite	fl_eruption_L1	file_name	id	confirmed	info_sc				
			0		Anchorage/FVAK21PAWU_1701121900.txt	311300	1	SATELLITE				
			-1		Anchorage/FVAK21PAWU_1701150130.txt	311300	1	HIMIWARI/GOES/SEISMICITY				
			0		Anchorage/FVAK21PAWU_1701181405.txt	311300	1	HIMIWARI/GOES/AVO/PILOT/REPORT/LIGH				
			0		Anchorage/FVAK21PAWU_1701181910.txt	311300	1	HIMAWARI/GOES				
			0		Anchorage/FVAK21PAWU_1701190140.txt	311300	1	GOES/MODEL/DATA				
			2		Anchorage/FVAK21PAWU_1701190635.txt	311300	1	GOES/POES/				
			0		Anchorage/FVAK21PAWU_1701201515.txt	311300	1	HIMAWARI/GOES/POES/AVO/PIREPS				
			0		Anchorage/FVAK21PAWU_1701201950.txt	311300	1	HIMAWARI/GOES/POES				
			2		Anchorage/FVAK21PAWU_1701210140.txt	311300	1	HIMIWARI/GOES/POES				
			0		Anchorage/FVAK21PAWU_1701221805.txt	311300	1	HIMAWARI/GOES				

vaac_total_line.json

Volcanic Dataset

```
{"colour_code": null, "fl_eruption_L2": null, "no_id_satelite": 0, "fl_eruption_L1": null, "file_name": "Darwin/2004/IDD41295.200412170409.txt", "issued_date": "20041217", "confirmed": 1, "info_source": "AVHRR/MODIS/USAF", "issued_time": "0408", "id": "252010", "vaac_obs": "Darwin", "referral": 0, "fl_fcst_18hr_L2": null, "fl_fcst_18hr_L1": null, "fl_fcst_12hr_L1": null, "fl_fcst_12hr_L2": null, "nxt_date": "NO_NXT_ADV", "report_id": "IDD41295", "name": " LANGILA", "fl_obs_cld_L1": null, "fl_fcst_6hr_L1": null, "fl_fcst_6hr_L2": null, "fl_obs_cld_L2": null, "nxt_time": "NO_NXT_ADV", "v_id": "252010"}
```

```
{"colour_code": null, "fl_eruption_L2": null, "no_id_satelite": 0, "fl_eruption_L1": null, "file_name": "Darwin/2004/IDD41295.200412170950.txt", "issued_date": "20041217", "confirmed": 1, "info_source": "AVHRR/MODIS/GOES9", "issued_time": "0949", "id": "252010", "vaac_obs": "Darwin", "referral": 0, "fl_fcst_18hr_L2": null, "fl_fcst_18hr_L1": null, "fl_fcst_12hr_L1": null, "fl_fcst_12hr_L2": null, "nxt_date": "NO_NXT_ADV", "report_id": "IDD41295", "name": " LANGILA", "fl_obs_cld_L1": null, "fl_fcst_6hr_L1": null, "fl_fcst_6hr_L2": null, "fl_obs_cld_L2": null, "nxt_time": "NO_NXT_ADV", "v_id": "252010"}
```


Metadata of the columns

- colour_code :
 - Green
 - Yellow
 - Orange
 - Red
 - None
- no_id_satellite:
 - Nonidentifiable form satellite (-1)
 - METEOROLOGICAL CLOUDS (1)
 - WX CLDS (2)
 - Dissipated (3)
 - Do not say (0) / we consider identified
- confirmed (information):
 - confirmed (1)
 - unconfirmed (-1)
 - Do not say (0)
- referral:
 - no issued time/date (1)
 - issued time/date (0)
- id : volcano id
- name: volcano_name
- vaac_obs: VAAC name
- report_id: report number
- fl_fcst_(6/12/18)hrs_L(1/2): Forecast of plume height & extend in 6/12/18 for L1/2
- fl_eruption_L(1/2): Eruption details for L1/2
- fl_obs_cld_L(1/2): Observed ash for L1/2
- issued_time: time issued the report
- issued_date: date issued the report
- nxt_time: next time for next report
- info_source: e.g. MTSAT-1R

Explore the volcanic dataset - ideas

- Analyse the number of reports that we have by year or by VAAC or by volcano
- Analyse the Flight Levels (eruption or observed)
 - Variables: fl_eruption_L1 ; fl_eruption_L2 ; fl_observed_L1; fl_observed_L2
- Analyse the colors (warning) of a volcano
- Analyse which is the volcano that erupted more during the last two years
- Analyse which is the volcano that erupted more across the entire dataset
- Analyse which volcanos appears in more than one VAAC

https://github.com/rosafilgueira/Seminar_MUIA/blob/main/lab_exercises/Volcano_Analysis/Volcanic%20data.ipynb

Exploring the Encyclopedia Britannica

https://github.com/EPCed/prace-spark-for-data-scientists/blob/master/lab_exercises/lab_rdd_dataframes_NLTK.ipynb

Exploring the Encyclopedia Britannica

https://github.com/rosafilgueira/Seminar_MUIA/blob/main/lab_exercises/Encyclopaedia_Analysis/lab_rdd_dataframes_NLTK.ipynb

NLTK Modules:

- `$python`
 - `>> import nltk`
 - `>> nltk.download("wordnet")`
 - `>> nltk.download("punct")`
 - `>> nltk.download('stopwords')`

Spacy:

You can also use **spacy** instead **NLTK**:

`$ pip install spacy --user`

Reimplement NLTK functions in the **notebook**:

`sent_tokenize_func`, `remove_stop_words_func`,

- Ideas:
 - Get the number of pages per year.
 - Get the number of words per year.
 - Frequency of topics per year:
 - Normalize results by number of words
 - Plot Frequencies

- Sports: golf, rugby, tennis, football

- Cities: Aberdeen, Dundee, Edinburgh, Glasgow, Stirling, Inverness, Perth

- Animal: cat, dog, kitten, puppy, goldfish, parrot

- Philosophers: Francis Hutcheson, David Hume, Adam Smith, Dugald Stewart, Thomas Reid

<https://medium.com/towards-artificial-intelligence/text-mining-in-python-steps-and-examples-78b3f8fd913b>

<https://www.dataquest.io/blog/tutorial-text-classification-in-python-using-spacy/>