

A Study of Significant Volcanic Eruptions

by Jaqueline Dias

Student Number: 501074392

Supervisor's name: Ashok Bhowmick, Ph. D

Submission Date: 14/02/2022



Table of Contents

Abstract	4
Introduction.....	6
Literature Review	6
The Volcanic Explosivity Index (VEI): An Estimate of Explosive Magnitude for Historical Volcanism.....	6
Long-Term Forecasting of Volcanic Explosivity	7
Anticipating Future Volcanic Explosivity Index (VEI) 7 Eruptions and Their Chilling Impacts	7
Correlations Between Earthquakes and Large Mud Volcano Eruptions.....	8
Volcanic Eruptions and Climate	8
Seasonality of Volcanic Eruptions.....	9
The Effect of Volcanic Eruptions on Global Precipitation	10
Climatic Response to High-Latitude Volcanic Eruptions	10
Dataset Description	11
Table 1 - Dataset 1: Eruptions	11
Figure 1: Distribution of Qualitative Features Eruption Dataset	13
Table 2 - Dataset 2: Volcano.....	14
Figure 2: Distribution of Qualitative Features Volcano Dataset.....	16
Approach.....	17

Chart 1: Flow Chart of Project Approach	17
Step 1: Selecting and Creating the Data Frame.....	17
Step 2: Cleaning the Data	18
Step 3: Analyzing the Data.....	18
Step 5: Presenting the Data	19
References	20

Abstract

Volcanoes and their eruptions have an impact on everyone's daily life whether they are aware of it or not. Besides the most evident impact resulting from a volcanic eruption, such as loss of life, change in the landscape, monetary loss and high recovery costs, there are long term effects that have been influencing life as we know for hundreds of years.

Volcanic eruptions have serious effects on climate change as their discharge contributes to lowering air quality. The pollution generated by the volcanic ashes and dust may linger for months after the eruptions, having a direct impact on water reserves, food availability, global travel and the lives of both humans and animals.

Most volcanoes are located between tectonic plates, and are characterized by being an opening on earth's surface allowing magma, ash and gases to escape. An eruption happens when gases and lava (magma that has reached earth's surface) are released from the volcano. This release can be, but not limited to, explosive eruptions. Most eruptions are of scale 0 to 2 on the Volcanic Explosivity Index (VEI). VEI is an index used to measure the explosiveness of volcanic eruptions, the index ranges from 0 to 8, with 8 being the most explosive volcanoes.

This project will study volcanic eruptions from 1020 to 2020, the sample datasets are available in the Kaggle webpage (see References). The focus of the prediction model will be eruptions considered significant, such eruptions are thus classified due to having caused fatalities,

moderate damage of approximately or superior to \$1 million, VEI index of 6 or greater and were associated with an earthquake of large magnitude or generated a tsunami.

Once merged, the dataset will have 37 independent variables and 6851 records. The data will be collected, prepared and cleaned to create the train and test datasets. The analysis will focus on different multiclass classification methods such as KNN and Decision trees to test and predict the Volcanic Explosivity Index of most significant volcanic eruptions. A second approach is proposed, that consists of a binary classification task where I will be using imbalanced classification to predict significant eruptions (VEI 6 to 8) and non-significant eruptions (VEI 0 to 5). Python is the selected tool for performing these analyses and the exploratory data analysis.

Introduction

Significant eruptions are low-probability, but high-consequence events, therefore knowledge of the frequencies of explosive eruptions is highly useful in a variety of volcano studies and in mitigating the impacts of such intense explosions on the environment and general life.

Furthermore, the objective of this study is to predict Volcanic Explosivity Index (VEI) as an attempt to prove whether it is possible to determine the volcano magnitude based on volcano characteristics. My interest in this is to help contribute to historical volcano data which is often known for being incomplete or poorly known for robust estimation of large events.

Literature Review

In this section, past studies related to volcanic explosivity magnitude and the factors that contribute to such events will be reviewed. It is valid to mention that “The main criterion for estimating VEI is bulk volume of pyroclastic deposits, but other criteria can be used when the volume is not well known.” (Newhall, Self and Robock 2018, p. 15). Eruptions considered significant are the ones that have caused fatalities, moderate damage of approximately or superior to \$1 million, VEI index of 6 or greater and were associated with an earthquake of large magnitude or generated a tsunami.

The Volcanic Explosivity Index (VEI): An Estimate of Explosive Magnitude for Historical Volcanism

This article dates from when VEI was first introduced and aims to present VEI as an alternative for quantitative estimate of eruption magnitude scale, which so far had been often attempted to be created but not satisfactorily succeeded. According to Newhall and Self (1982), the main challenge in creating a scale able to identify the magnitude of an eruption was the lack of common parameters for historical volcanic eruptions recorded so far, given that the same parameters would have to be analyzed for each volcanic activity, however historic data was often incomplete.

To reach its goal, this study assigned VEI over 8000 historic and prehistoric eruptions. The study concludes that VEI is a useful measure to compare the explosivity of historical eruptions and a

measure able to assist in evaluating the completeness of historical records, contributing to the validation of studies using those records. It is important to note that throughout the years not only what is introduced in this article has been proved true, but VEI has become largely used in the geologic field as a quantitative measure of volcanic explosivity magnitude. For those reasons contributing to the classification of past eruptions is proposed by my study.

Long-Term Forecasting of Volcanic Explosivity

The purpose of the above study was to determine whether an extended interval between eruptions influences the VEI of a volcano's next eruption. The study stated that probabilistic forecasting of eruption size is one of the major scientific challenges regarding volcanoes, and risk analysis so far rely on evidence-based approaches and experts recommendation to determine the size of the next eruption.

Using data from volcanoes in Indonesia since the early 19th century, Bebbington (2014) considers characterizing VEI distribution by conduit state, open or closed, using a Bayesian multilevel modeling approach. The study concludes that while the size of the previous eruption has no effect on the size of the following one, there is a significant probability that the VEI of the next eruption from closed conduit volcanoes increases with increasing repose length. Open conduit volcanoes, however, show no evidence of increase in VEI with increase in repose length, which agrees with what Gregorio and Carmada study suggests that "The greater the amount of surplus of magma within the feeding system, the higher is the eruptive probability." (Gregorio and Camarda, 2016, p. 1), based on these conclusions, repose length will not be used on the model phase of my analysis.

Anticipating Future Volcanic Explosivity Index (VEI) 7 Eruptions and Their Chilling Impacts

The purpose of the above study is to review historical and geological evidence of high-end subduction related volcanic eruptions, more specifically those of VEI 7. Newhall, Self and Robock reinforce the importance of establishing parameters to identify candidates for future VEI 7 eruptions as well as discuss the challenges for long and short-range forecasting of such unlikely events. However unlikely, such events have great repercussions in the entire world.

The study continues on stating that modern times have not yet dealt with the results of VEI 7 eruptions, as they are likely to happen 1-2 times per thousand years. Eruptions of such magnitude can produce at least 100 km³ of ash and large masses of sulfur, carbon dioxide, and other volcanic gasses that will circle earth's surface many times, causing air pollution, climate change,

such as sulfate aerosol from tropical eruptions. Their residue from secondary explosions that can occur even a year after the main event, can go as high as reaching commercial flights.

The study then suggests 6 criteria that will help in long-range forecasting of VEI 7 eruptions, and following with short-range (immediate) forecasting that can help making response plans in case of eventuality. The study concludes reinforcing that VEI 7 eruptions are not unthinkable, and not only populations within a range of volcanoes are at great risk, but the risk is also global and could even pose a threat to peace between nations, therefore governments and industry are suggested to use new technologies to anticipate most new vulnerabilities and how they might be diminished.

Correlations Between Earthquakes and Large Mud Volcano Eruptions

This study aims to examine the correlation between large earthquakes and the triggering of eruptions in mud volcanoes, specifically with eruptions whose moment magnitude was greater than 5 and volcanoes who were situated within a 100 km of these earthquakes. This study also aims to examine the potential increase of the number of mud volcano eruptions after these large earthquakes. Eruptions, or more specifically, magma movement may cause earthquakes and a volcano prompt to erupt may be triggered by earthquakes, therefore this is a very important topic to be included in this review.

The above-mentioned study uses a relatively accurate catalog of mud volcano eruptions as its primary data source, and while acknowledging the potential for typographical error, endeavors to use the most accurate subset of this data source, using data between the years 1960-2001. This is compared against a list of the seismic intensities of earthquakes within the dates of mud volcano eruptions. Mellors et al. (2007), measure the number of eruptions in a calendar year following an earthquake, comparing this to the number of eruptions in the calendar years before the earthquake.

The study concludes that while large earthquakes can trigger mud volcanoes eruptions (predominantly through seismic liquefaction), this is not always the case and points to the fact that other factors must play an important role in the triggering process. The study also concludes that there is a weak correlation between large earthquakes and delayed onset of mud volcano activity.

Volcanic Eruptions and Climate

This study intended to observe the effects of volcanic activity on the earth's climate, focusing primarily on the injection of anthropogenic material into the atmosphere from these eruptions.

A number of large volcanic eruptions have been speculated to have been a cause for abnormal weather patterns throughout history, evident in unexpected seasonal temperatures and agricultural crop failures.

The researcher first summarizes the constituents of volcanic input to the atmosphere (predominantly magmatic material, along with a number of different gases such as H₂O, N₂, CO₂ and SO₂) and their effect on both short and long wavelength radiation.

They then review new ice core samples as an insight to past volcanic activity and compare these to past ice core analyses. Next, they review the effect of these eruptions on local diurnal cycles as well as the impact of these eruptions on climate on a decadal and century scale time range and their contributions to the Little Ice Age by El Nino events.

Robock (2000), finds that the injection of these volcanic gases can have a multi-year effect on the climate. A large part of this is due to forward and back scattering of incident solar radiation, resulting in a net cooling effect at earth's surface. The study does conclude however that while there is potential for volcanic activity to have large temporal scale effects on climate changes, any El Nino events observed around times of large volcanic activity were coincidental.

Seasonality of Volcanic Eruptions

This study aims to analyze seasonal patterns in eruption rates of all sizes on global and regional scales by using the Smithsonian catalog of eruptions recorded over the last 300 years. The Smithsonian catalog is the same catalog used in my study, however different sample data is considered in Mason et al. (2004) study.

The study notes that there are clear seasonal patterns in eruptions of smaller magnitude, which had not been noticed previously as most studies analyzed larger eruptions. Therefore, smaller or nonexplosive eruptions of VEI 0 to 2 are more likely to present seasonal volcanic patterns. The study analyzes the influence of the lunar tidal phase on volcanic activity and finds no conclusive relationship between the events.

Next, this study tries to show regional patterns by seasonal fluctuations in sea level, using coastal tide gauge data and fluctuations in regional atmospheric pressure by examining observations of the start date of single eruptions and significant eruption sequences, then grouping them into twelve bins (the bins aim to correspond to the months of the year). Mason et al. (2004), conclude that volcanic activity of smaller magnitude shows significant statistical seasonality along the Pacific Ring of Fire and at some individual volcanoes.

The Effect of Volcanic Eruptions on Global Precipitation

This study sets out to examine the effect that large, low latitude volcanic eruptions have on global precipitation. The study uses 18 historical volcanic eruptions between the years 1400 and 2000 and analyses global precipitation trends after these eruptions using the Global Historical Climatology Networks precipitation data set. The study also carries out some climate modeling techniques using the coupled climate model HadCM3. Using simulation runs from this modeling, they compare the results to actual observations and quantify the agreement between these using detection analysis.

The study aims to examine the difference in timing between land and ocean precipitation response, as well as the extent of specific features of land precipitation responses to volcanic eruptions. The study concludes that in their modeling, there was an expected precipitation response for around 5 years over oceanic regions along with a temperature response over the same timeline. This is compared to a 3 year precipitation response over land and this precipitation response reacted faster than the temperature response. Comparing these models to observational data, they observed that the model response was significant for the boreal cold season, however it was only marginally significant for the boreal warm season.

Climatic Response to High-Latitude Volcanic Eruptions

The goal of this study is to gain some insight into the climatic response caused by the eruption of high latitude volcanic eruptions; volcano latitude is one of the parameters I will be analyzing in the modeling phase of my study, hence the importance of reviewing this article. The study uses two volcanic eruptions in its modeling, both are high latitude eruptions, however they vary in the optical depth caused by the strength of their eruptions. Oman et al. (2005) leveraged the Goddard Institute for Space Studies ModelE GCM modeling technique in its experimentation.

The study concludes that the main impact from these eruptions appears to be radiation caused by the injection of aerosols (such as SO₂) into the atmosphere. This anthropogenic material leads to higher heating levels in the lower stratosphere. It was noted that sea level pressures decreased for the following winter for the weaker eruption compared to the next 2 winters for the larger eruption. A trend of warmer temperatures in monsoon regions with reduced cloud coverage was also noted, and even more evident when compared with the results of the more powerful volcanic eruption.

Dataset Description

The data frame used in this project will be a combination of two datasets, the first being about volcano eruptions, which shows the incidence of eruptions per volcano throughout time, and the second dataset has characteristics of volcanoes, such as major and minor rock, tectonic settings, elevation, and so on. Both datasets will be joined by volcano number, which is the index used to identify the volcanoes in both sets. The datasets are available at the Kaggle website, link is available in the references at the end of this document.

The initial procedures and analysis of the dataset used in this study can be accessed on the repository created for this project on my GitHub account at the following link: <https://github.com/JaqueD/Capstone-Project>

Next, all variables in both datasets will be listed and quantitative variables will have mean, median, standard deviation, min and max values included in the respective tables and distribution histograms charts will follow.

Table 1 - Dataset 1: Eruptions

Feature Name	Description	Mean	Median	Standard Deviation	Min	Max
Volcano Number	Volcano identification number	300284.37	290050	52321.19	210010	600000
Volcano Name	Name attributed to each volcano					
Eruption Number	Eruption identification number	15666.91	15650.5	3297.61	10001	22355
Eruption Category	Confirmation status of eruption					
Area of Activity	Area of activity					
VEI	Volcanic Explosivity Index	1.95	2	1.16	0	7

Start Year	Year when eruption started	622.85	1847	2482.17	-11345	2020
Start Month	Month when eruption started	3.45	1	4.07	0	12
Start Day	Day when eruption started	7.01	0	9.65	0	31
Evidence Method Dating	Evidence for dating volcanic eruption					
End Year	Year when eruption ended	1917.33	1957	157.65	-475	2020
End Month	Month when eruption ended	6.22	6	3.69	0	12
End Day	Day when eruption ended	13.32	15	9.83	0	31
Latitude	Volcano's coordinate for north-south position	16.87	17.60	30.76	-77.53	85.61
Longitude	Volcano's coordinate for east-west position	31.57	55.71	115.25	-179.97	179.58

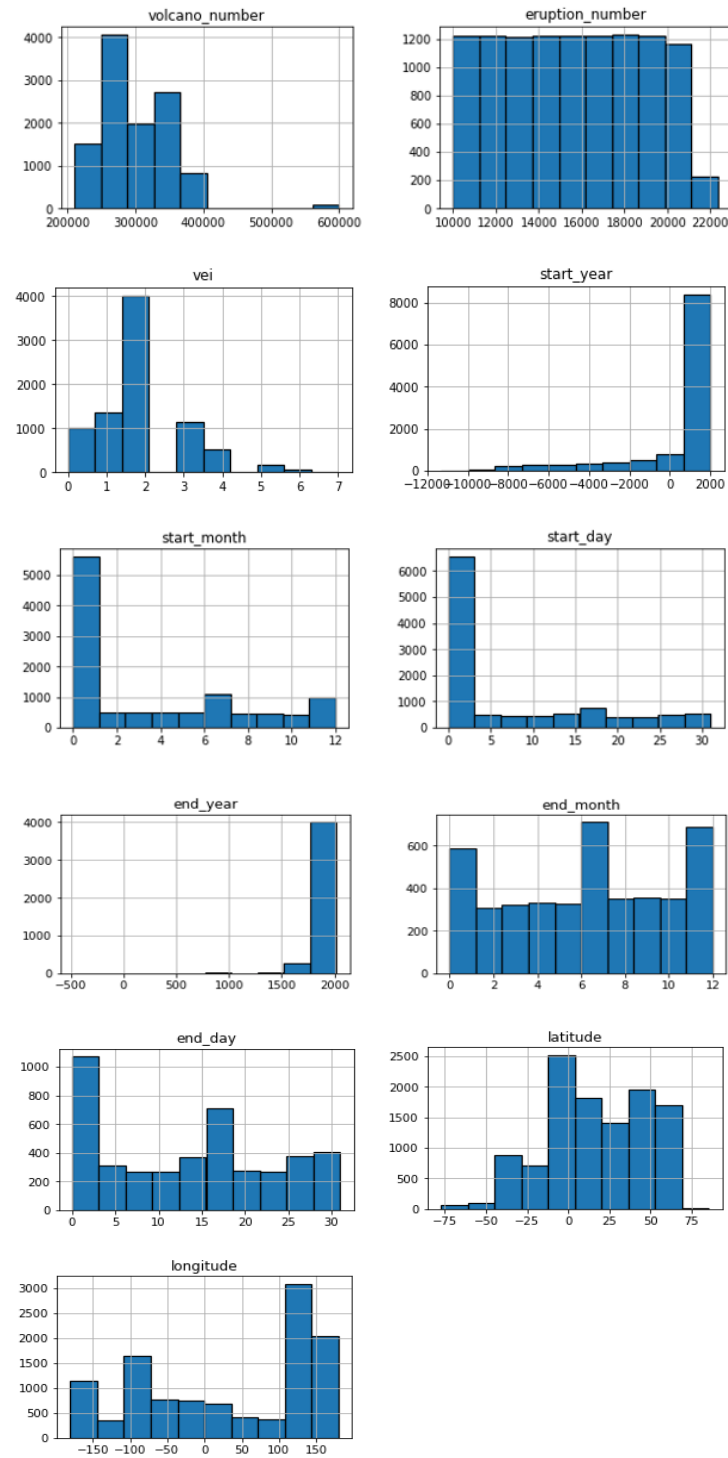
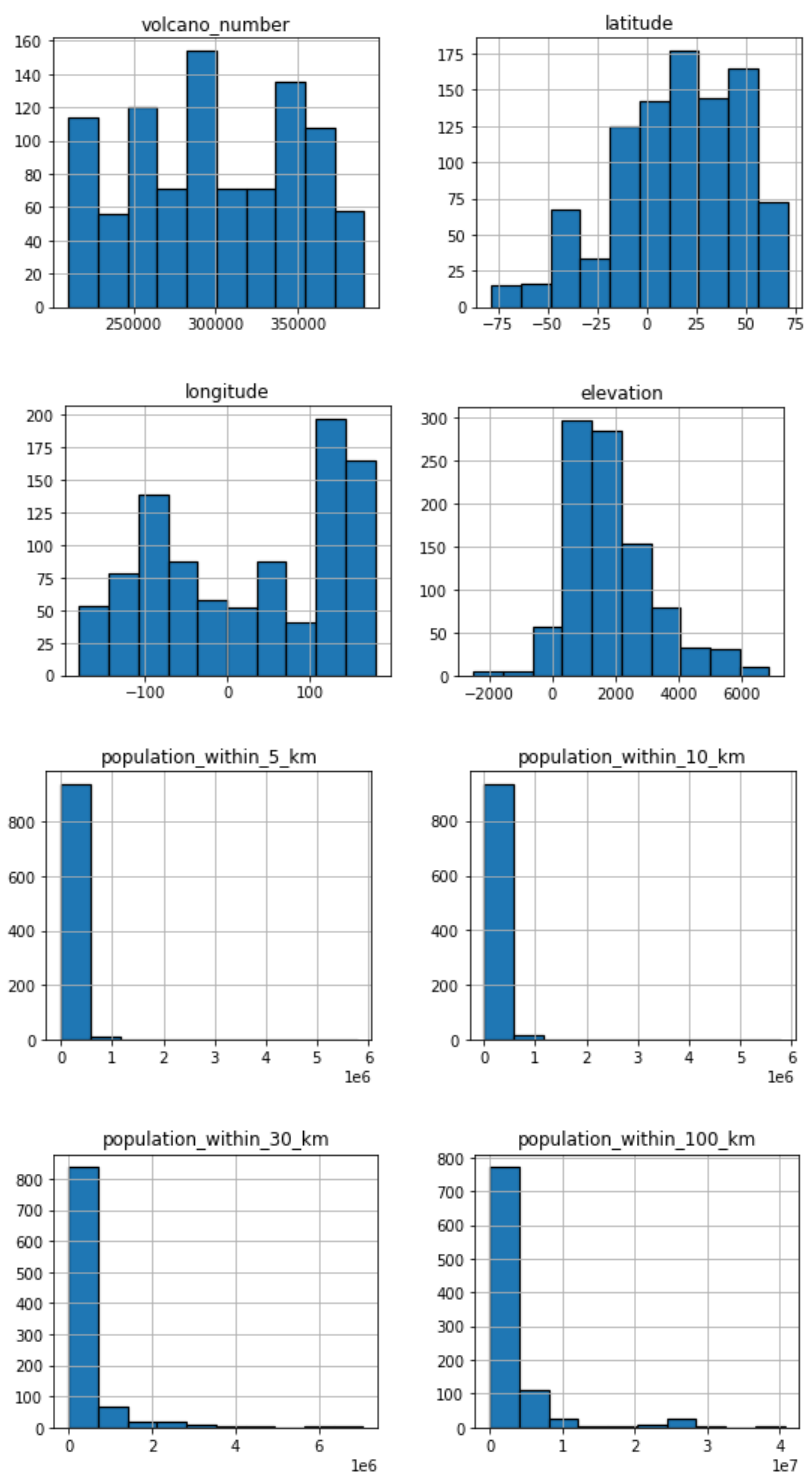
Figure 1: Distribution of Qualitative Features Eruption Dataset

Table 2 - Dataset 2: Volcano

Feature Name	Description	Mean	Median	Standard Deviation	Min	Max
Volcano Number	Volcano identification number	298585.33	300055.50	49792.66	210010	390829
Volcano Name	Name attributed to each volcano					
Primary Volcano Type	Primary volcano type					
Last Eruption Year	Year when last eruption occurred					
Country	Country of occurrence					
Region	Region of occurrence					
Subregion	Subregion of occurrence					
Latitude	Volcano's coordinate for north-south position	14.98	14.51	31.58	-78.50	71.08
Longitude	Volcano's coordinate for east-west position	23.54	36.39	109.85	-179.97	179.58
Elevation	Volcano elevation	1867.03	1622.50	1401.55	-2500	6879
Tectonic Settings	Plate-tectonic setting of volcanoes					
Evidence Category	Category of evidence					

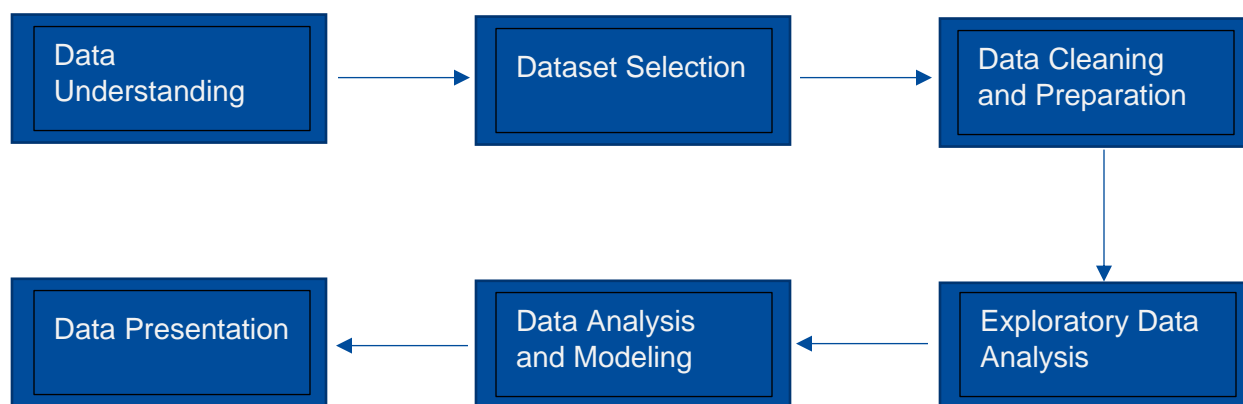
Major Rock 1	Rock type					
Major Rock 2	Rock type					
Major Rock 3	Rock type					
Major Rock 4	Rock type					
Major Rock 5	Rock type					
Minor Rock 1	Rock type					
Minor Rock 2	Rock type					
Minor Rock 3	Rock type					
Minor Rock 4	Rock type					
Minor Rock 5	Rock type					
Population Within 5 km	Sum of population within 5 km	4.79e+4	2.95e+2	2.99e+5	0	5.78e+6
Population Within 10 km	Sum of population within 10 km	6.12e+4	1.63e+3	3.02e+5	0	5.78e+6
Population Within 30 km	Sum of population within 30 km	3.04e+5	1.39e+4	7.35e+5	0	7.07e+6
Population Within 100 km	Sum of population within 100 km	2.73e+6	3.55e+5	5.69e+6	0	4.06e+7

Figure 2: Distribution of Qualitative Features Volcano Dataset

Approach

In this topic I will describe the procedures that will be used in the preparation of the dataset, cleaning the data, analysis and final presentation of the complete study and its results.

Chart 1: Flow Chart of Project Approach



Step 1: Selecting and Creating the Data Frame

The two datasets used in this project share the same source and both were exported as CSV files. The datasets will be combined into a data frame by volcano number using Python. Volcano number is a variable common in both datasets used as a unique volcano identifier. Dataset 1: Eruptions, has records of volcanic eruptions that have happened throughout time for each volcano, it has 11178 entries and 15 variables. Dataset 2: Volcano, registers volcanoes characteristics and records the population living within 5, 10, 30 and 100 km from the volcano location, this dataset has 958 entries and 26 variables.

Once merged the data frame will have 9559 entries and 40 variables. Volcanoes without records of eruptions are not included in the final data frame, therefore, records in this study rely on the data of volcanoes which have had one or several eruptions throughout time.

Categorical variables such as major rock 1, 2, 3, 4 and 5, minor rock 1, 2, 3, 4 and 5, as well as tectonic settings will be mapped into numerical value and assign to a new variable in order to be used as predictor variables in the modeling part of this project, which will aim to predict the VEI of significant volcanoes eruptions, therefore VEI is the dependent variable in this analysis. How

the dependent variable will be treated is still to be determined, however two approaches are proposed, first as a multiclass variable considering all VEI instances (0 to 8), or a binary variable using VEI 0 to 5 as non-significant and VEI 6 to 8 as significant.

Step 2: Cleaning the Data

Once merged there are some features that will be duplicated, such as volcano name, latitude and longitude, in that case the duplicates will be deleted and there will be 37 independent variables left in the dataset. Another point to be considered is the time period for the data frame. Considering the present records, they date from 11345 BC to 2020, that is too large a period for the purpose of this analysis, as advised by my supervisor Ashok Bhowmick, Ph. D. For that reason, the dataset will be filtered between 1020 to 2020, and will contain a total of 6851 records.

When filtered to a more recent period most of the variables in the dataset contain complete data entries for each volcanic eruption. However, some features have missing data points and to decide what treatment to use I will consult my supervisor in order to determine the best course of action in each variable case, should it be filled with mean, mode or dropped if necessary.

Step 3: Analyzing the Data

At the analysis part of this project, Python is once more the selected tool for performing the exploratory data analysis and modeling phase of the data.

There are two approaches that I will be consulting with my supervisor to decide which one will suit best the purpose of this project's analysis. The first approach will focus on different multiclass classification methods such as KNN and Decision trees to test and predict the VEI of most significant volcanic eruptions, using VEI index 0 to 8 as a dependent variable. The second approach will be a binary classification task where I will be using imbalanced classification to predict significant eruptions (VEI 6 to 8) and non-significant eruptions (VEI 0 to 5), this approach seems relevant given that the majority of volcanic eruptions in the dataset belong to the non-significant class and a minority of records belong to the significant eruption class.

I will verify whether the dataset benefits from scaling methods and analyze the correlation between the variables. Once the relevant variables for predicting the dependent variables have been determined, I will create the predictive model.

Step 5: Presenting the Data

I will be using Tableau to create some of the visualizations for the PowerPoint presentation, some of the charts created in Python for the exploratory data analysis will also be included. The presentation shall display the results obtained from the prediction models implemented as well as summarize the methods implemented and conclusions drawn.

References

- Chris Newhall, Stephen Self, The Volcanic Explosivity Index (VEI): an estimate of explosive magnitude for historical volcanism, *Journal of Geophysical Research: Oceans*, Volume 87 (Issue C2), February 1982, Pages 1231-1238, <https://doi.org/10.1029/JC087iC02p01231>.
- M. S. Bebbington, Long-term forecasting of volcanic explosivity, *Geophysical Journal International*, Volume 197 (Issue 3), June 2014, Pages 1500-1515, <https://doi.org/10.1093/gji/ggu078>.
- Sofia De Gregorio, Marco Camarda, A novel approach to estimate the eruptive potential and probability in open conduit volcanoes, *Scientific Reports*, Volume 6, Article number: 30471, July 2016, Pages 1-7, <https://doi.org/10.1038/srep30471>
- Chris Newhall, Stephen Self, Alan Robock, Anticipating future Volcanic Explosivity Index (VEI) 7 eruptions and their chilling impacts, *Geosphere*, Volume 14 (Issue 2), February 2018, Pages 572-603, <https://doi.org/10.1130/GES01513.1>
- R. Mellors, D. Kilb, A. Aliyev, A. Gasanov, G. Yetirmishli, Correlations between earthquakes and large mud volcano eruptions, *Journal of Geophysical Research: Solid Earth*, Volume 112 (Issue B4), April 2007, <https://doi.org/10.1029/2006JB004489>.
- Alan Robock, Volcanic eruptions and climate, *Reviews of Geophysics*, Volume 38 (Issue 2), May 2000, Pages 191-219, <https://doi.org/10.1029/1998RG000054>.
- B. G. Mason, D. M. Pyle, W. B. Dade, T. Jupp, Seasonality of volcanic eruptions, *Journal of Geophysical Research: Solid Earth*, Volume 109 (Issue B4), April 2004, <https://doi.org/10.1029/2002JB002293>.
- Carley E. Iles, Gabriele C. Hegerl, Andrew P. Schurer, Xuebin Zhang, The effect of volcanic eruptions on global precipitation, *Journal of Geophysical Research: Atmospheres*, Volume 118 (Issue 16), August 2013, Pages 8770-8786, <https://doi.org/10.1002/jgrd.50678>.

Luke Oman, Alan Robock, Georgiy Stenchikov, Gavin A. Schmidt, Reto Ruedy, Climatic response to high-latitude volcanic eruptions, *Journal of Geophysical Research: Atmospheres*, Volume 110 (Issue D13), July 2005, <https://doi.org/10.1029/2004JD005487>.

NCEI/WDS Global Significant Volcanic Eruptions Database, 4360 BC to Present. (2001, March 23). National Centers for Environmental Information (NCEI). Retrieved January 20, 2022, from <https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ngdc.mgg.hazards:G10147>

Jesse Mostipak. Volcano Eruptions. Kaggle, Retrieved January 25, 2022, from <https://www.kaggle.com/jessemostipak/volcano-eruptions?select=volcano.csv>

Volcano Hazards Program Glossary - VEI. (n.d.). USGS. Retrieved January 20, 2022, from <https://volcanoes.usgs.gov/vsc/glossary/vei.html>