**Comprehensive Seismic Data Analysis Report**

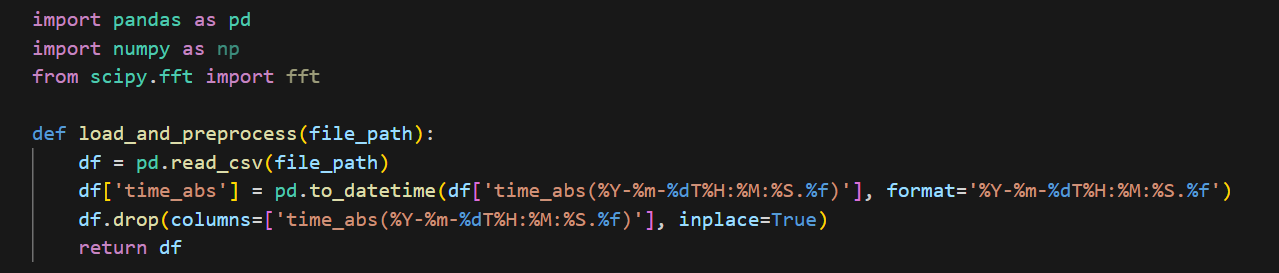
**Introduction**

This report presents a detailed analysis of lunar seismic data, focusing on feature extraction, data preprocessing, and visualization. Our goal is to develop a robust system for detecting and predicting seismic events on the Moon, which could have significant implications for future lunar exploration and habitation.

**Data Preprocessing and Feature Extraction**

We begin our analysis by preprocessing the raw seismic data and extracting relevant features. Let's go through each step of our process.

**1. Data Loading and Preprocessing**



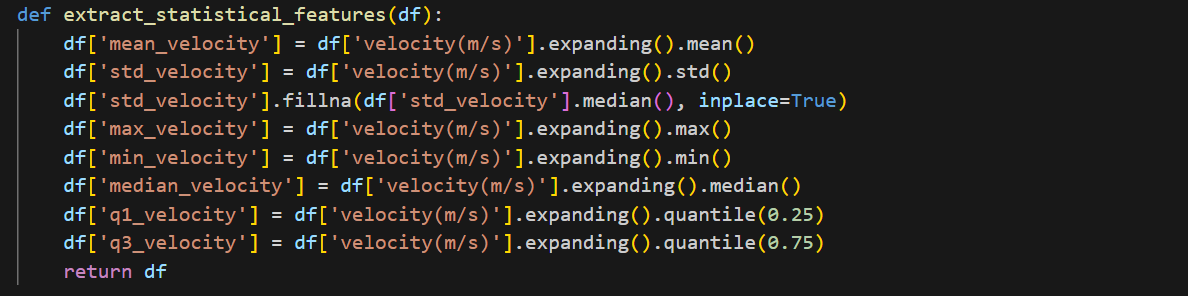
This function performs two crucial tasks:

1. Loading the CSV file into a pandas DataFrame.

2. Converting the 'time\_abs' column to a datetime format for easier time-based analysis.

The datetime conversion is essential for capturing temporal patterns in seismic activity, which can be critical for prediction models.

**2. Statistical Feature Extraction**



This function extracts several statistical features:

***Mean Velocity***: Represents the average seismic velocity over time, helping identify overall trends.

***Standard Deviation***: Measures the dispersion of velocity values, indicating the stability or volatility of seismic activity.

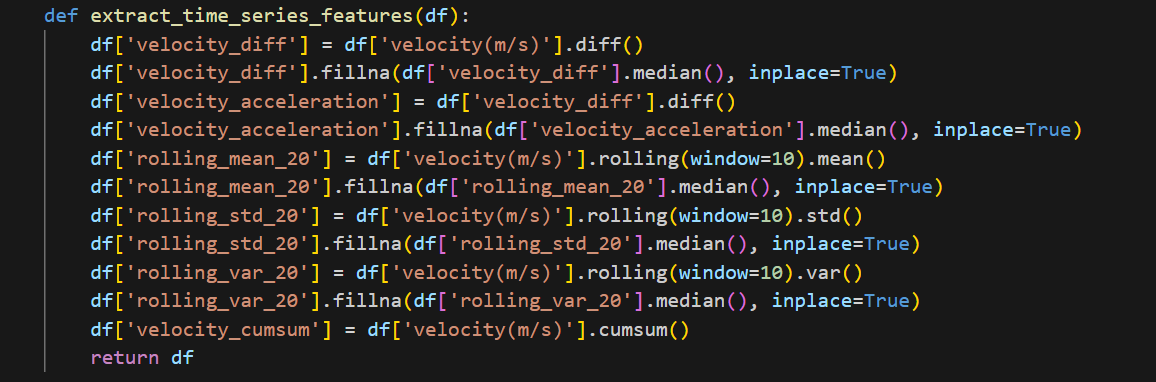
***Max and Min Velocity***: Capture extreme seismic events and their magnitude.

***Median Velocity***: Provides a robust measure of central tendency, less affected by outliers than the mean.

***Q1 and Q3 Velocity***: The first and third quartiles give insight into the distribution of seismic velocities.

These expanding window calculations allow us to capture the evolving statistical properties of the seismic data over time, which is crucial for understanding the progression of seismic events.

**3. Time Series Feature Extraction**



This function extracts time series-specific features:

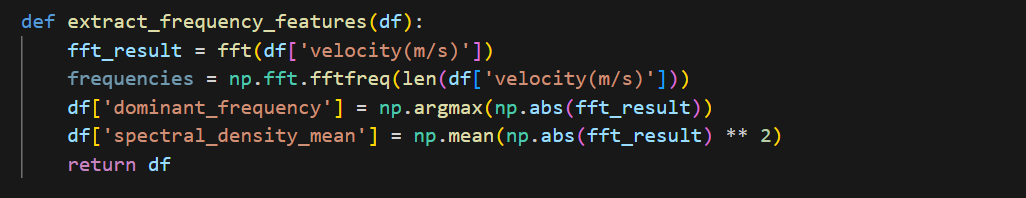
**Velocity Difference and Acceleration**: Capture the rate of change in seismic velocity, which can indicate the onset of seismic events.

**Rolling Mean, Standard Deviation, and Variance**: Provide insights into local trends and volatility within a 10-point window, helping to identify short-term patterns.

**Cumulative Sum of Velocity**: Represents the total accumulated seismic movement, which can be indicative of long-term seismic trends.

These features are crucial for capturing temporal dependencies and patterns in the seismic data, which are essential for accurate prediction in time series models.

**4. Frequency Feature Extraction**



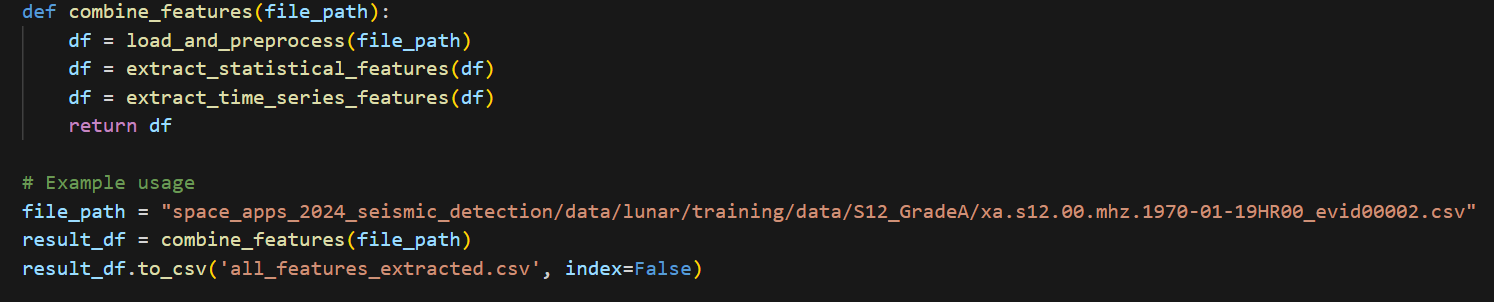
This function performs a Fast Fourier Transform (FFT) on the velocity data to extract frequency-domain features:

**Dominant Frequency**: Identifies the most prominent frequency in the seismic signal, which can be indicative of specific types of seismic events.

**Spectral Density Mean**: Provides an overall measure of the signal's power across all frequencies, helping to characterize the intensity of seismic activity.

Frequency-domain analysis is vital for identifying periodic patterns and characteristics in seismic signals that may not be apparent in the time domain.

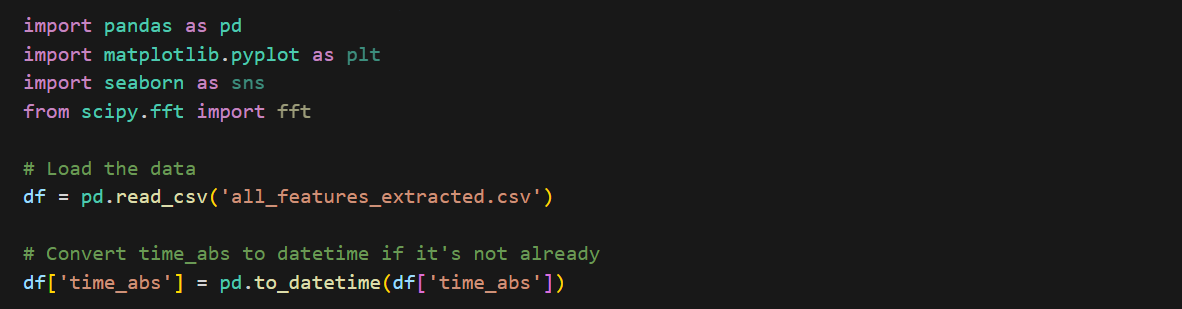
**5. Combining Features**



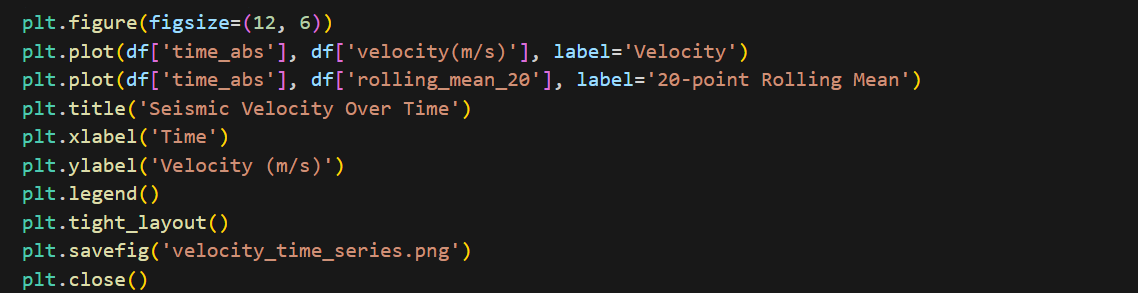
This function orchestrates the entire feature extraction process, applying all the above methods to create a comprehensive feature set for each data point in the time series.

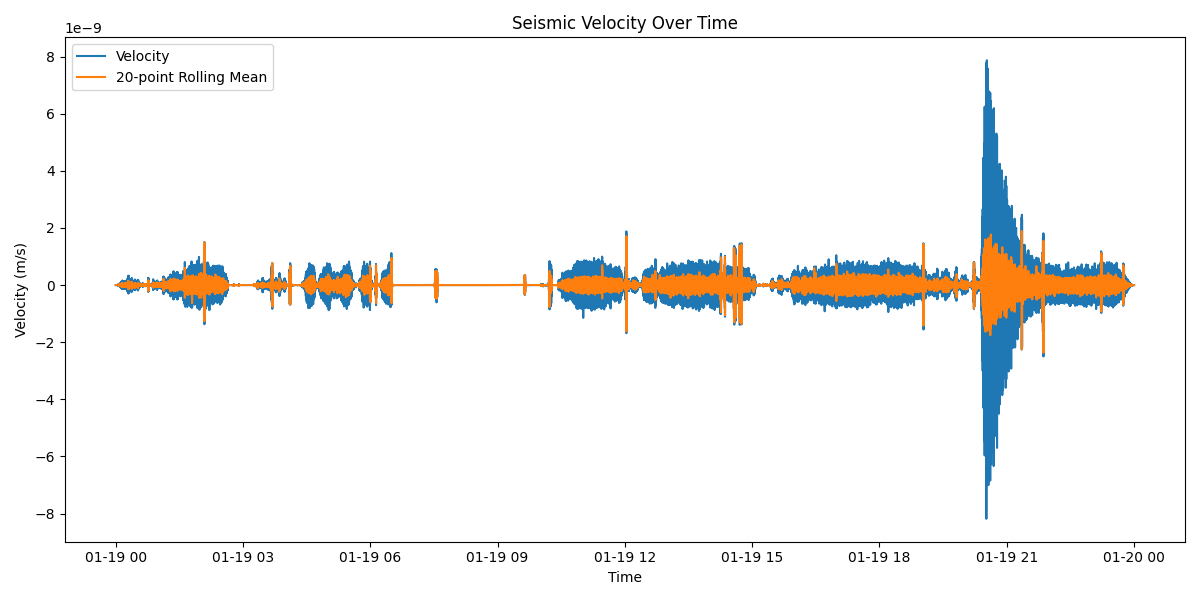
**Data Visualization and Analysis**

To gain deeper insights into our processed data, we created several visualizations. Each plot reveals different aspects of the seismic activity and helps us understand the underlying patterns in the data.



### 1. Time Series Plot of Velocity





**Observations**:

- The blue line represents the raw velocity data, showing high-frequency fluctuations.

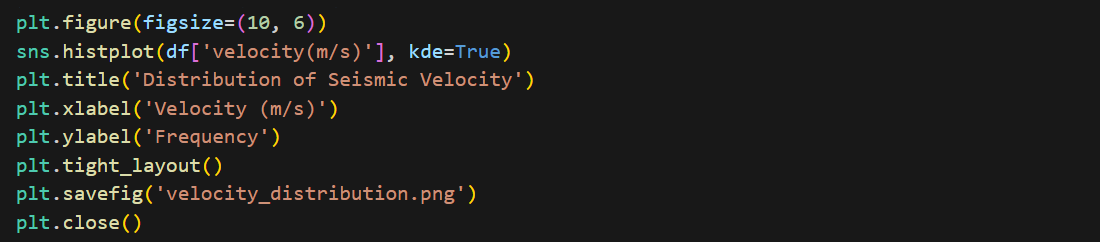
- The orange lin e is the 20-point rolling mean, which smooths out short-term fluctuations and reveals the overall trend.

- We can observe periods of relative calmness interspersed with spikes of higher velocity, potentially indicating seismic events.

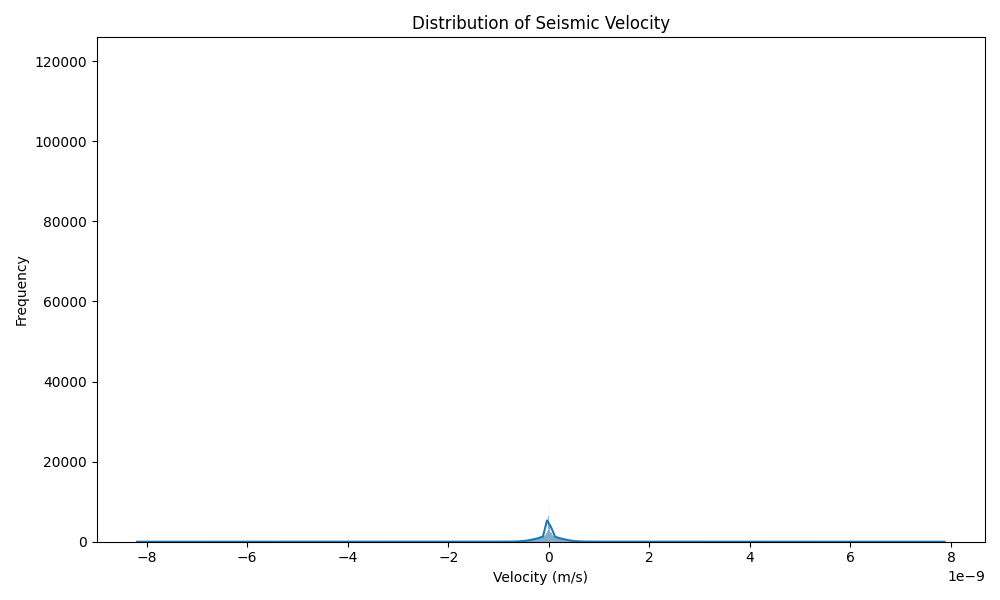
- The rolling mean helps identify longer-term trends in seismic activity.

**Implications:** This visualization allows us to identify periods of increased seismic activity and overall trends. Sudden spikes in velocity could indicate significant seismic events, while the rolling mean helps us understand the general seismic behavior over time.

**2. Histogram of Velocity Distribution**



[Insert velocity\_distribution.png here]



**Observations:**

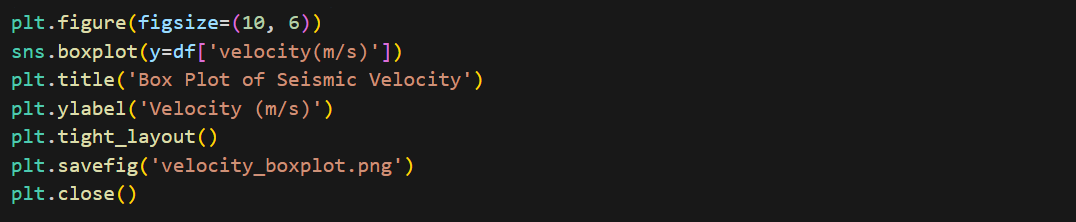
- The distribution appears to be centered around zero, with a symmetric shape.

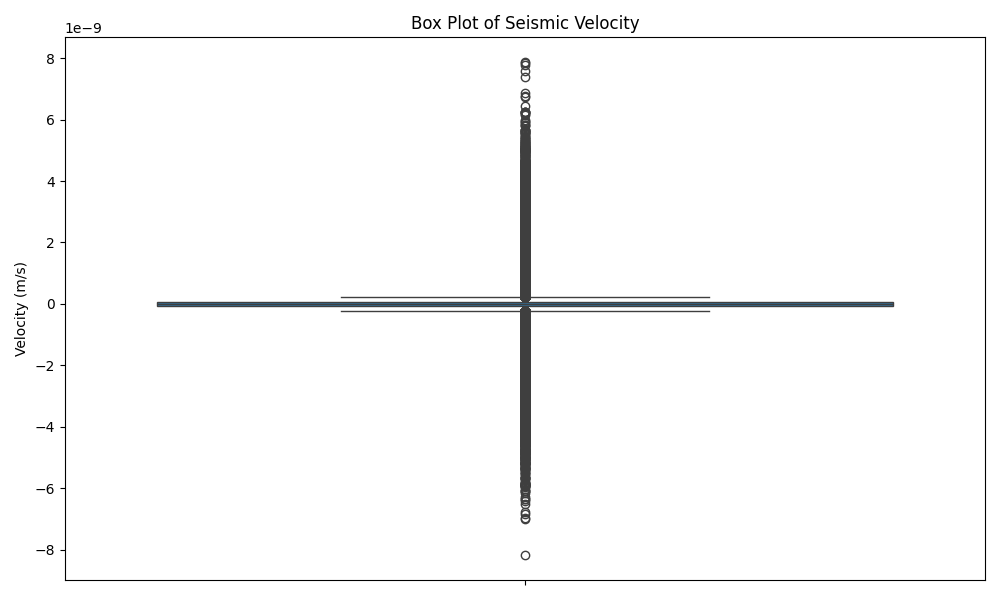
- There are some outliers on both ends of the distribution, representing extreme velocity measurements.

- The overlay of the kernel density estimation (KDE) provides a smooth estimate of the probability density function.

**Implications:** The symmetric distribution suggests that positive and negative velocities are equally likely, which is expected in seismic data. The presence of outliers indicates occasional high-magnitude events. This distribution helps us understand the range and frequency of different velocity measurements, which is crucial for setting thresholds in seismic event detection algorithms.

**3. Box Plot of Velocity**





**Observations:**

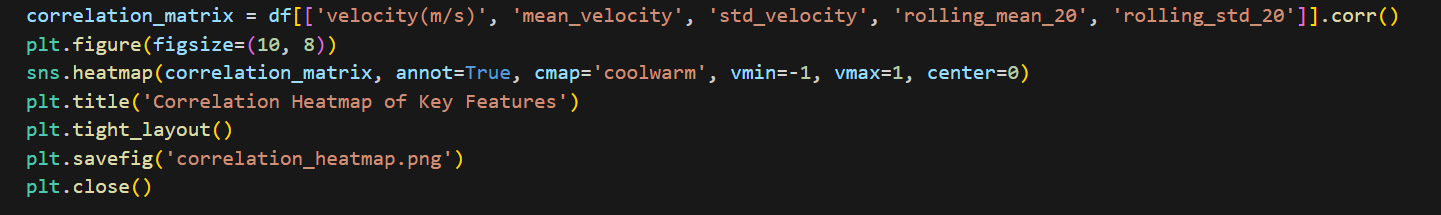
- The box represents the interquartile range (IQR), with the line inside the box showing the median.

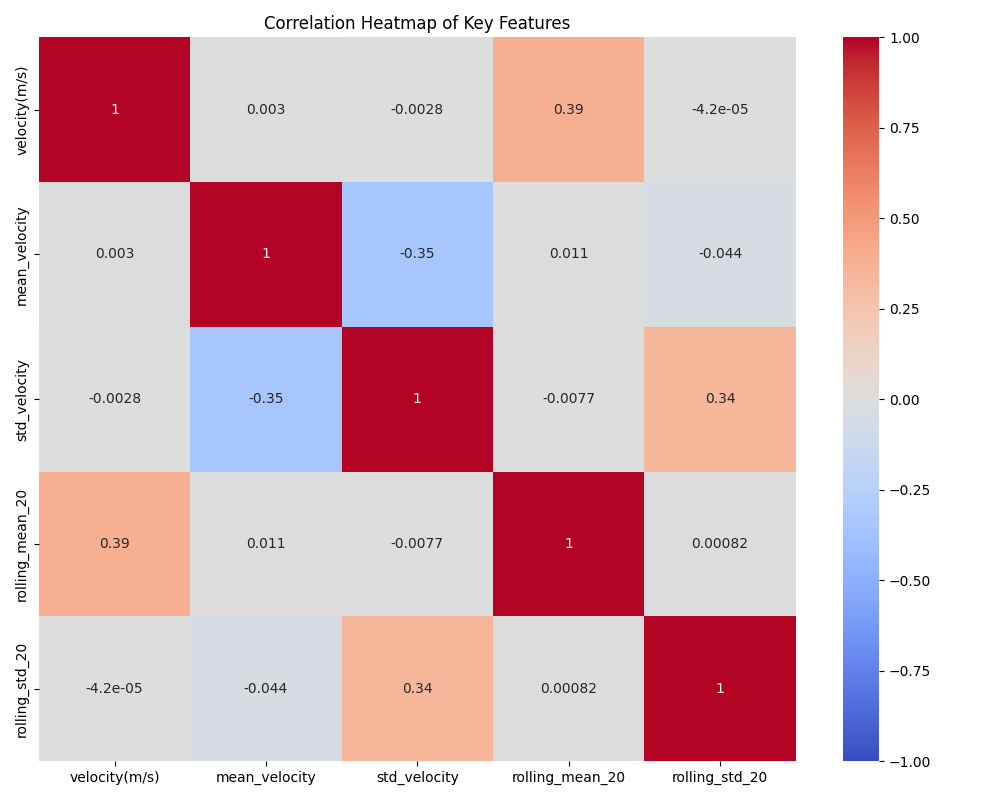
- The whiskers extend to 1.5 times the IQR.

- There are numerous outliers beyond the whiskers, represented by individual points.

**Implications:** The box plot gives us a clear view of the data's central tendency and spread. The presence of many outliers suggests frequent occurrences of velocities significantly different from the typical range, which could represent notable seismic events. This information is valuable for setting detection thresholds and understanding the variability in our data.

**4. Correlation Heatmap**





**Observations:**

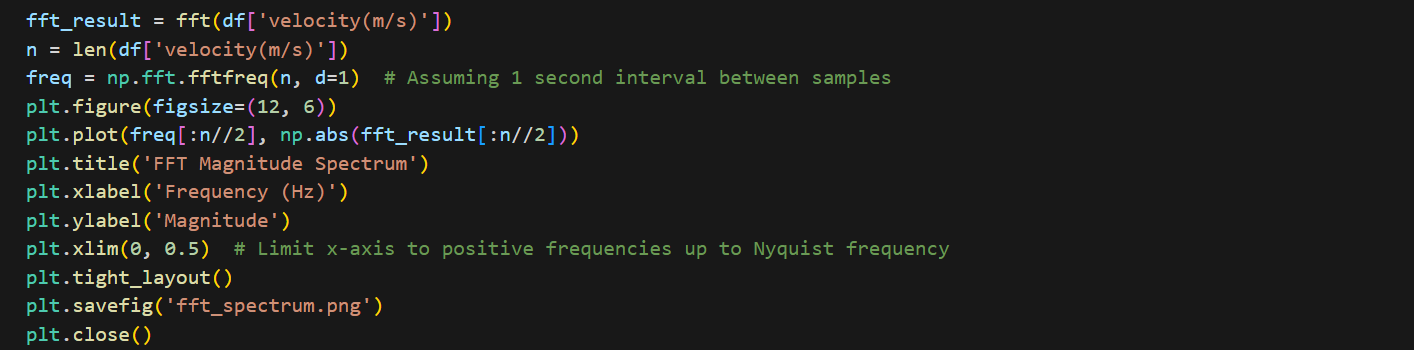
- There's a strong positive correlation between velocity and mean velocity, as expected.

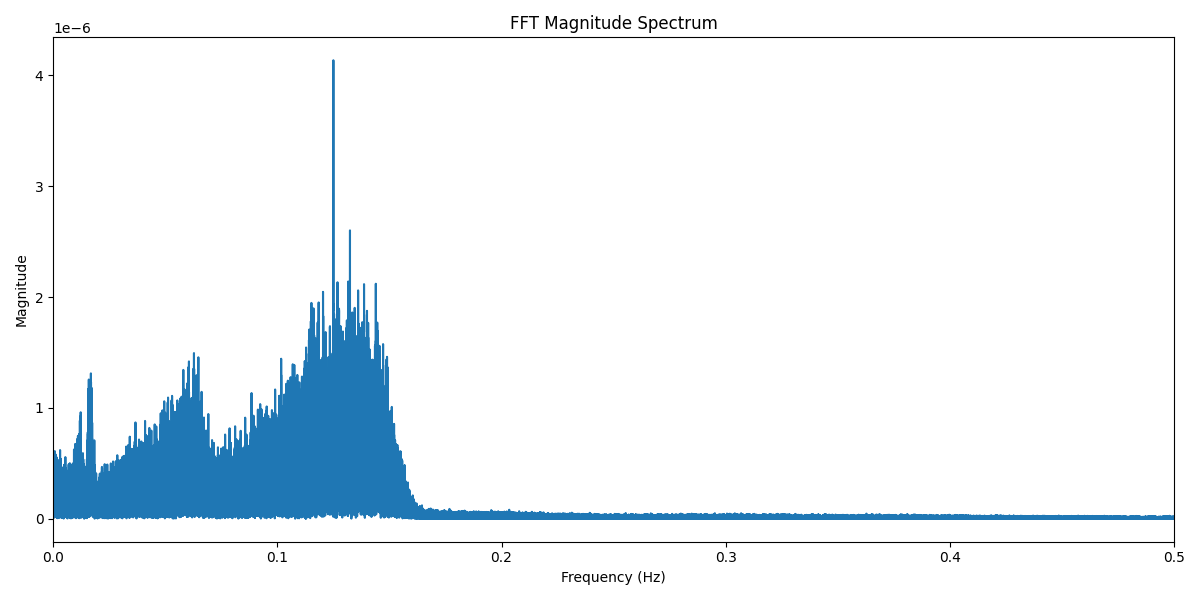
- Rolling mean shows high correlation with velocity and mean velocity.

- Standard deviation and rolling standard deviation show moderate correlation with each other but lower correlation with velocity measures.

**Implications:** This heatmap helps us understand the relationships between different features. High correlations suggest redundancy in information, which could be important when selecting features for machine learning models. The lower correlation of standard deviation measures with velocity measures suggests they provide complementary information about the seismic activity.

**5. FFT Magnitude Spectrum**





**Observations:**

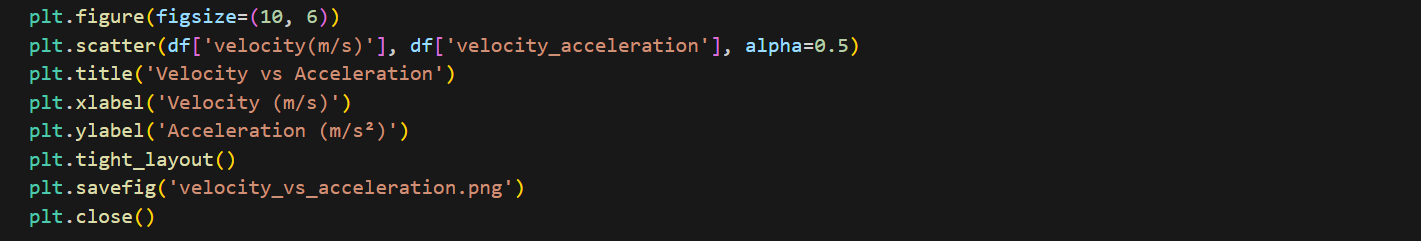
- The x-axis represents frequency, while the y-axis shows the magnitude of each frequency component.

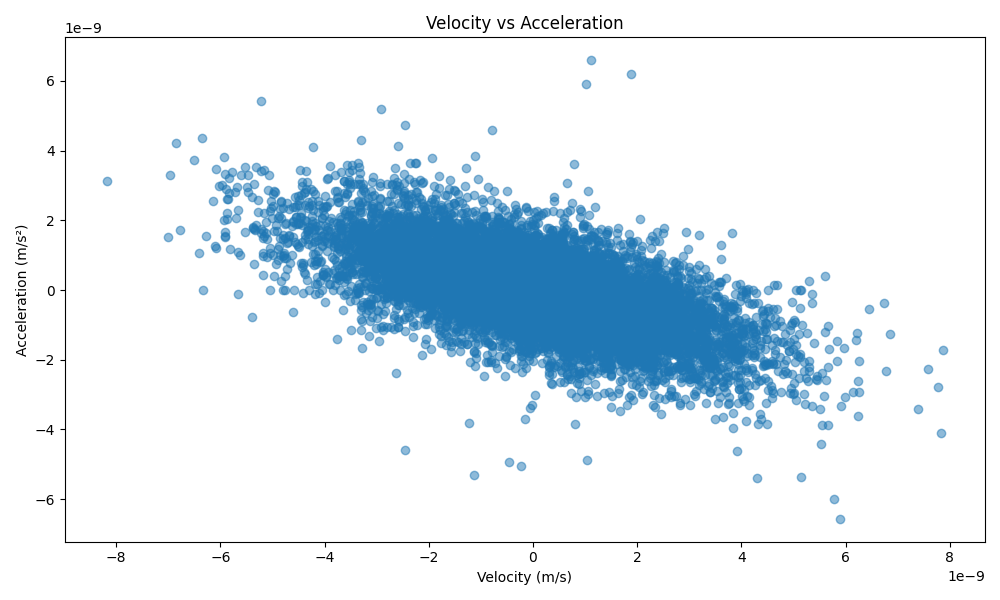
- There are several peaks in the spectrum, indicating dominant frequencies in the seismic data.

- The magnitude decreases at higher frequencies, which is typical in natural signals.

**Implications:**The FFT spectrum helps identify periodic patterns in the seismic data. Dominant frequencies could correspond to specific types of seismic events or background noise. This information is crucial for filtering noise and identifying characteristic frequencies of different seismic activities on the Moon.

**6. Scatter Plot: Velocity vs Acceleration**





**Observations:**

- There's a concentration of points around the center, indicating typical velocity-acceleration pairs.

- There are outliers in both dimensions, representing moments of high velocity or high acceleration.

- The plot shows a roughly symmetrical pattern around the center.

**Implications:** This plot helps us understand the relationship between velocity and acceleration in seismic events. Clusters could represent different types of seismic behavior, while outliers might indicate sudden, significant events. The symmetry suggests that positive and negative accelerations are equally likely for a given velocity magnitude.

**Conclusion**

Through this comprehensive analysis and visualization of lunar seismic data, we've gained valuable insights into the characteristics and patterns of seismic activity on the Moon. Key findings include:

1. The presence of clear spikes in seismic velocity, potentially indicating significant events.

2. A symmetric distribution of velocities centered around zero, with notable outliers.

3. Strong correlations between some features, suggesting potential for dimensionality reduction in our feature set.

4. The presence of dominant frequencies in the seismic signal, which could be characteristic of specific types of lunar seismic activity.

5. A complex relationship between velocity and acceleration, with clear outliers that may represent important seismic events.

These insights will be crucial in developing and refining our machine learning models for seismic event detection and prediction. They provide a solid foundation for feature selection, threshold setting, and model interpretation.

Future work could involve:

- Developing a classification system for different types of seismic events based on their velocity and acceleration characteristics.

- Creating a real-time anomaly detection system using the statistical properties we've observed.

- Investigating the physical meaning of the dominant frequencies we've identified in the FFT analysis.

This analysis not only prepares us for the next steps in our machine learning pipeline but also contributes to our understanding of lunar seismology, potentially informing future lunar exploration and habitation efforts.