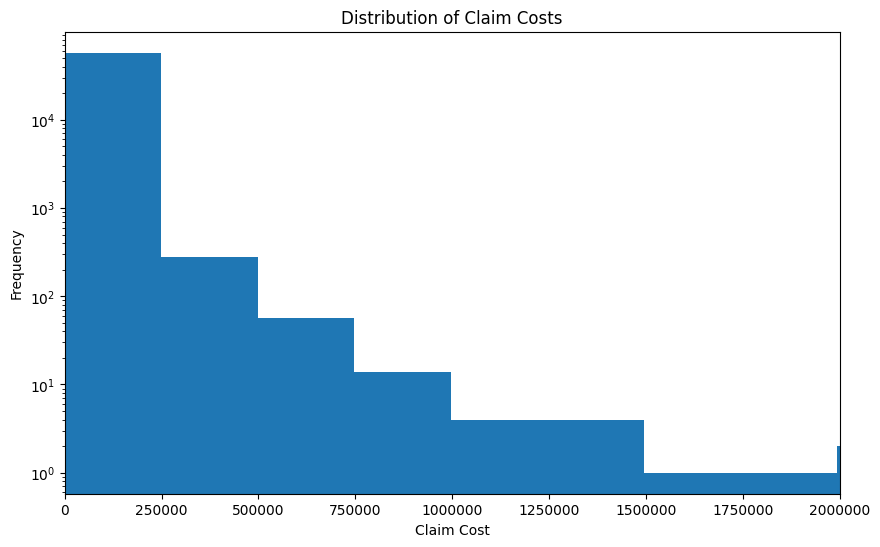
Univariate Analysis :

1. Checking distribution of Claim Cost

A graph of a distribution of costs

Description automatically generated

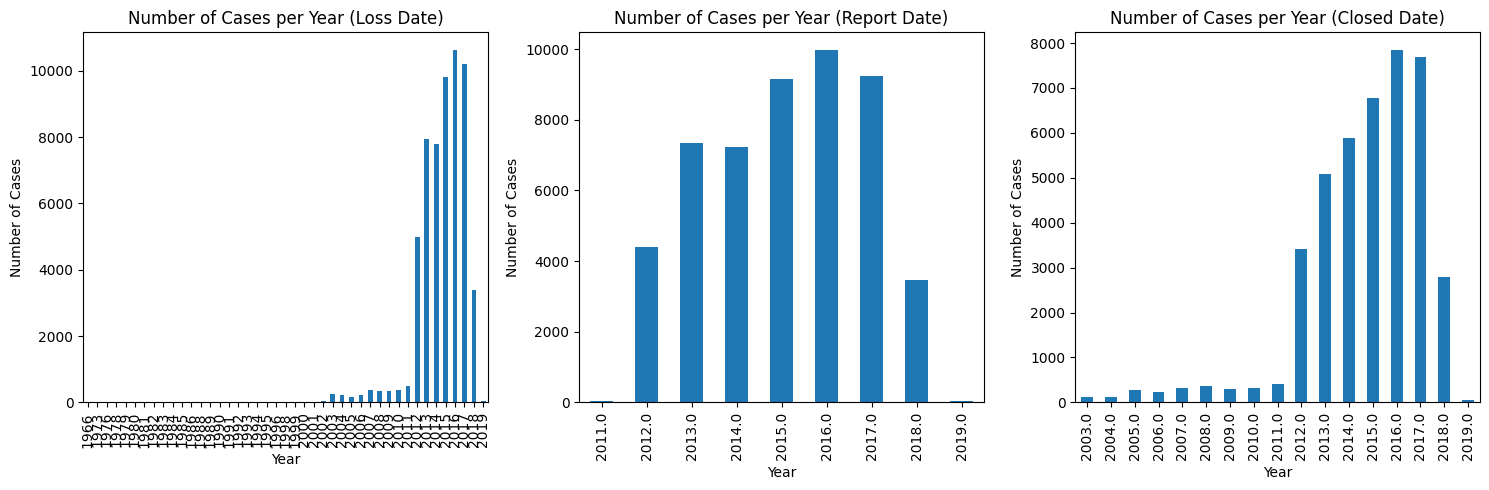
Zooming into claim costs upto 2 million



- The majority of claim costs are concentrated in the lower range (0 to 250,000) as evident from the high frequency of bars in this range

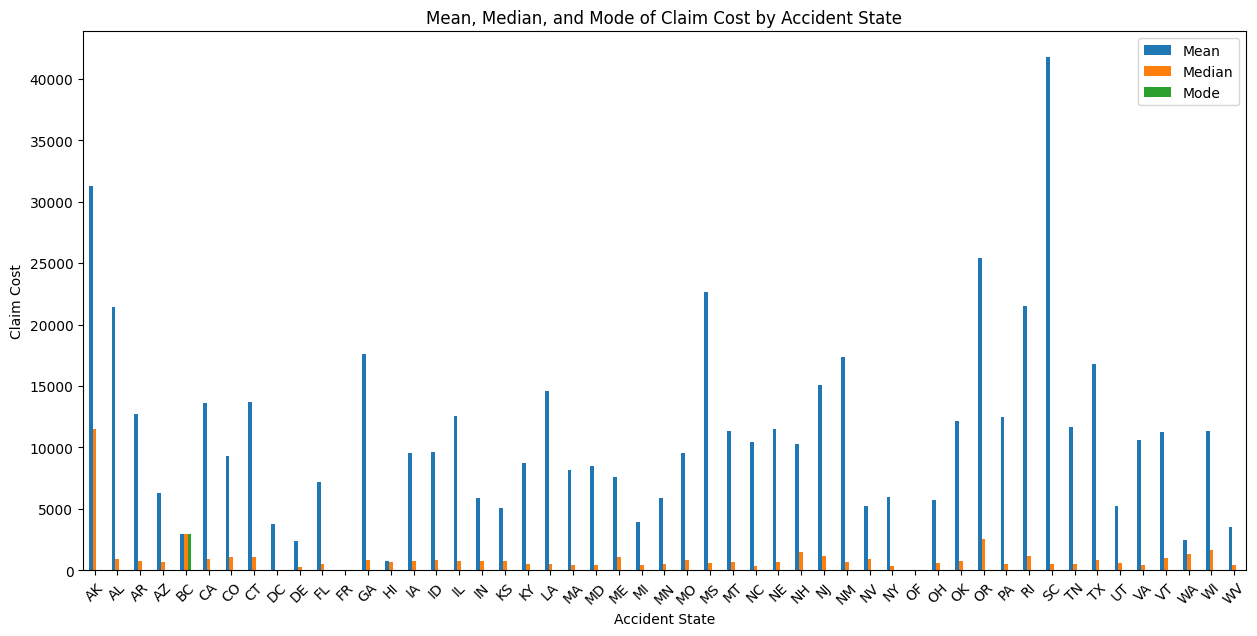
- Claims between 250,000 and 2 million are relatively sparse compared to the lower ranges. There are very few claims in the higher bins, indicating that such high-cost claims are rare.

1. Claim cost distribution across dates



Data before 2012 Loss Year can be pruned as we see volume of data from 2012 onwards.

1. Mean, Median, and Mode of Claim Cost by Accident State



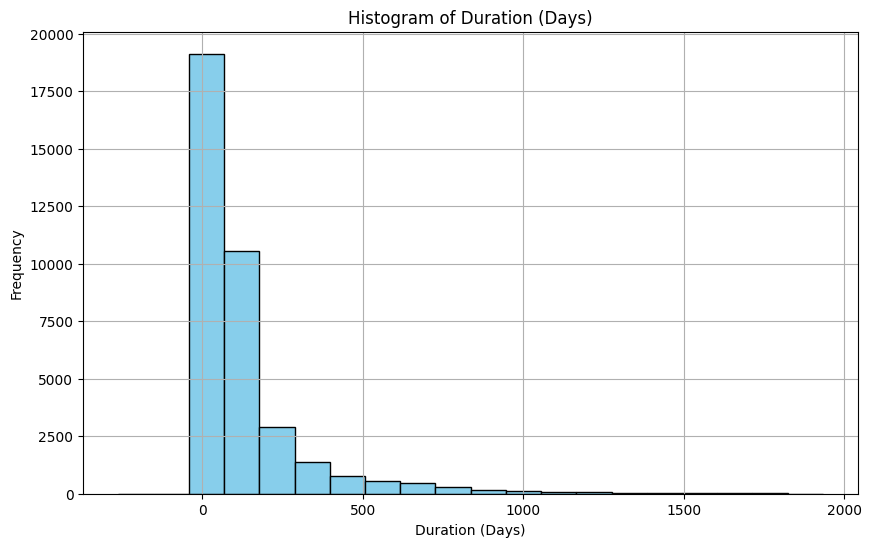
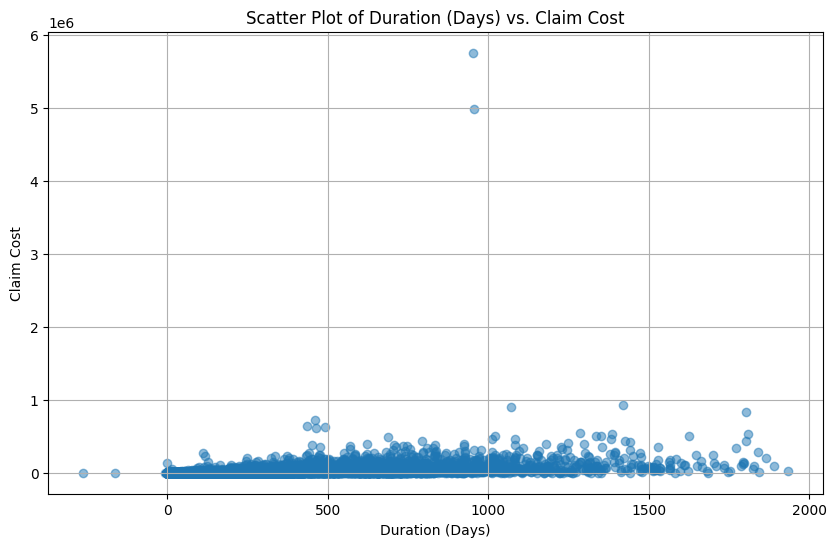
1. **High Cost States**: AK and SC have significantly higher mean claim costs, indicating high expenses.

3. **Outliers Impact**: Large mean-median gaps suggest outliers skewing average costs in all states except BC.

4. **Policy Adjustments**: Insurers might adjust premiums based on state-specific claim cost trends.

5. **Resource Focus**: Allocate resources to high-cost states for better fraud detection and claims management.

1. Impact of Duration of Claim (Gap between Report Date and Closed Date) on Claim Cost



Visualizing post outliers removal :-

A graph of blue dots

Description automatically generated

- The majority of the claim costs are clustered within a specific range of durations, indicating that most claims are resolved within a certain period.

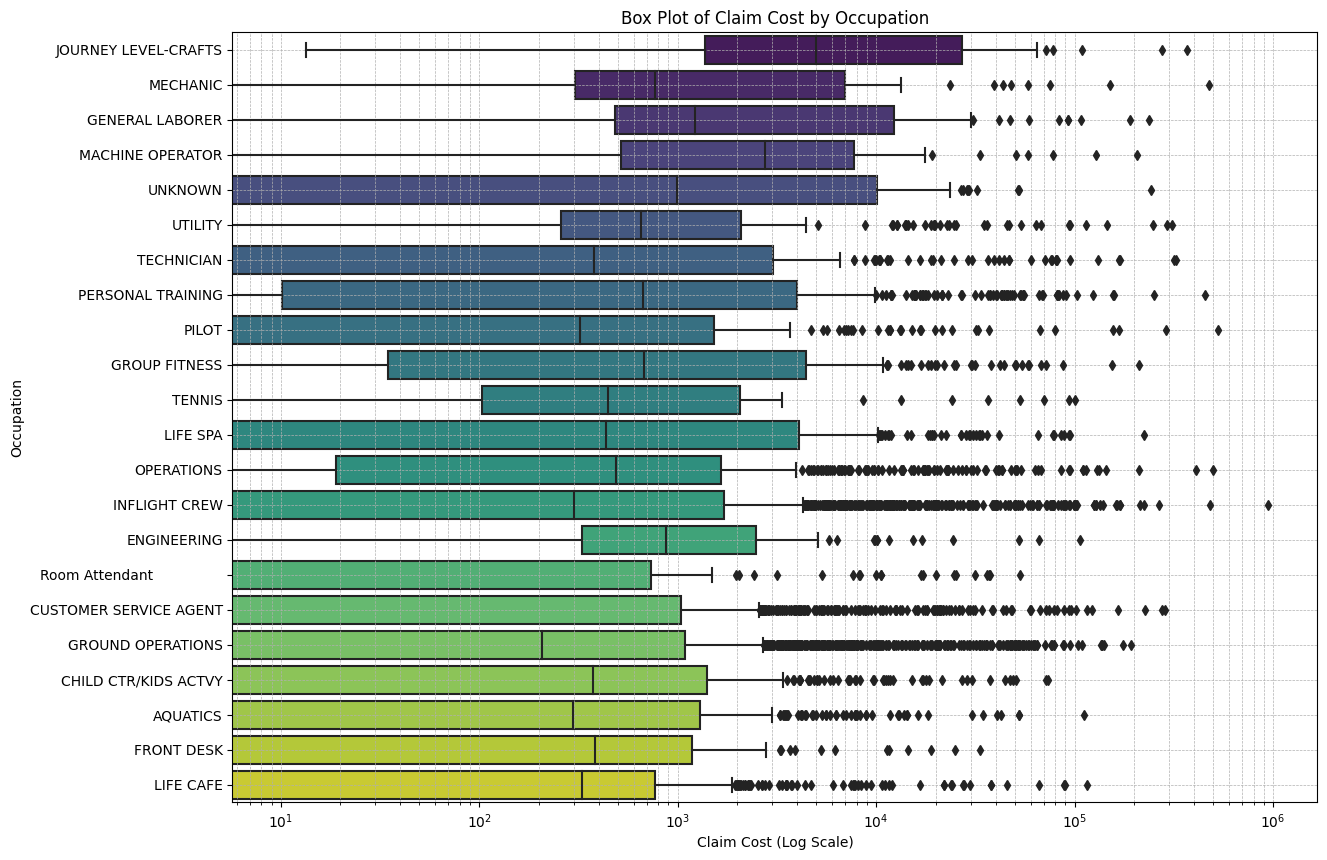
- There are fewer data points as the duration increases, suggesting that prolonged claim processes are less common.

**- There doesn't appear to be a strong linear relationship between the duration of the claim process and the claim cost. This implies that the duration alone may not be a significant predictor of the claim cost.**

1. **Occupation vs Claim Cost**

***Note* : Occupation data is available for about 30% of the total records. Hence the following is not a true representation of the entire dataset.**

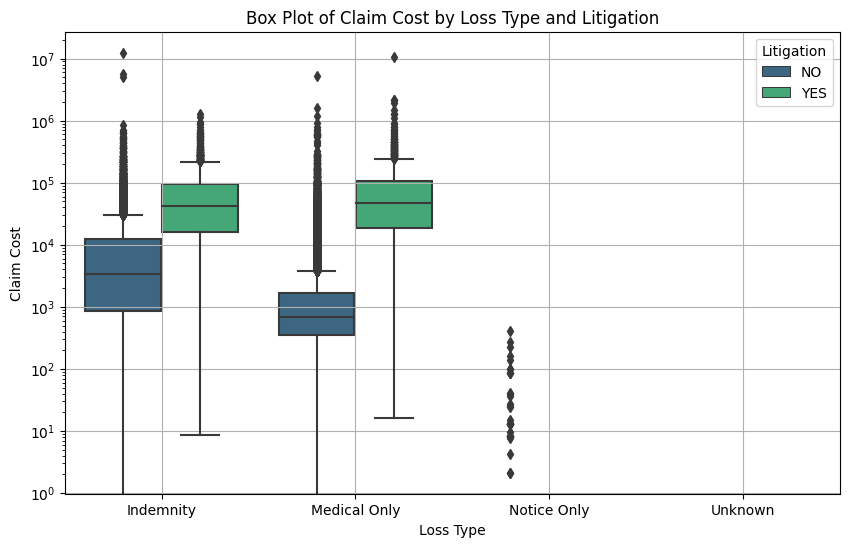
**However, investigating the relationship between claim costs and occupation could highlight the importance of capturing this information more comprehensively in future data collection efforts.**

****

**- Occupations - JOURNEY LEVEL-CRAFTS | MECHANIC | GENERAL LABORER | MACHINE OPERATOR | UTILITY | ENGINEERING have higher median claim costs compared to others**

**Bivariate Analysis :**

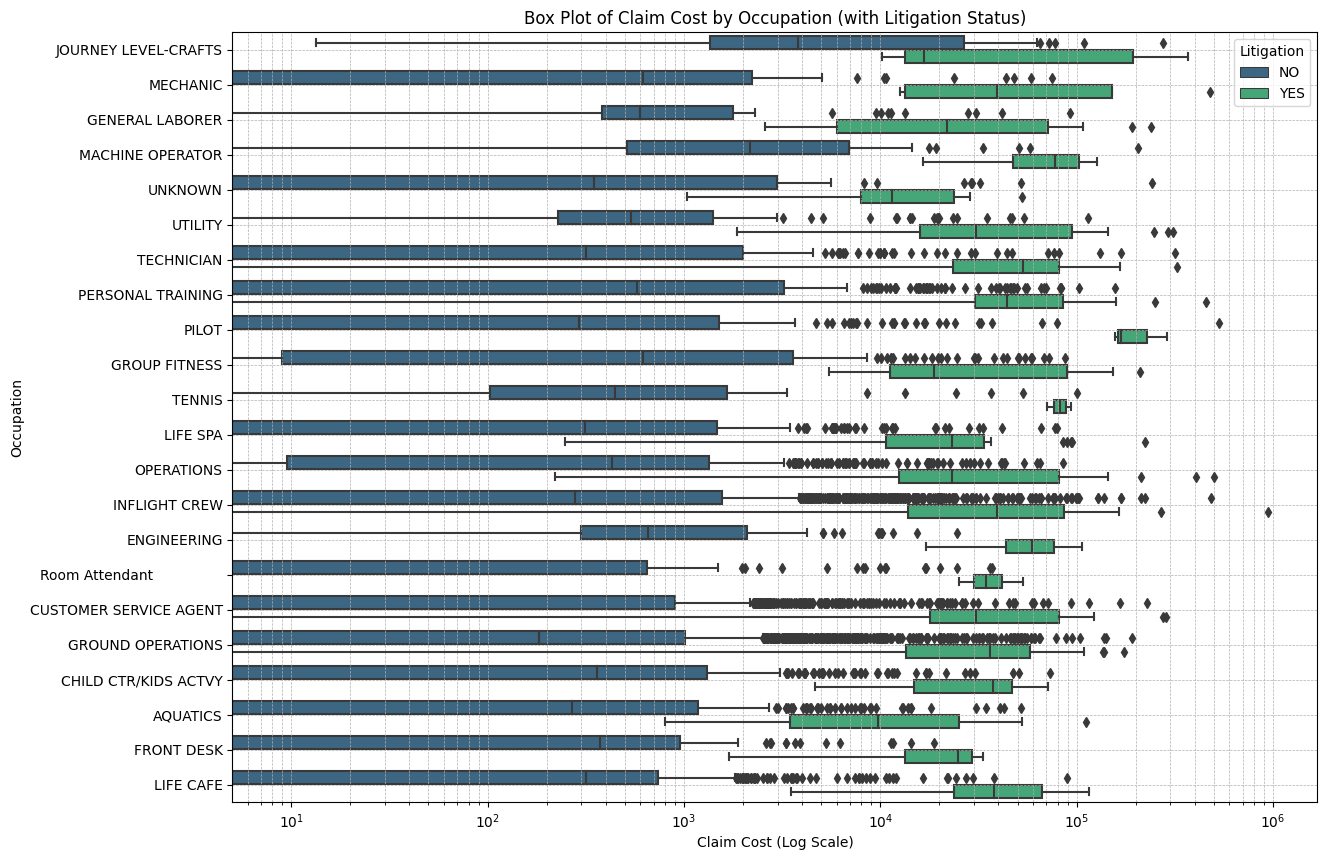
1. **Claim Cost vs [Loss Type and Litigation]**

****

- 'Indemnity' has higher median claim costs compared to 'Medical Only'.

- Claims with litigation have higher median claim costs compared to those without litigation.

1. **Claim Cost vs [Occupation and Litigation]**

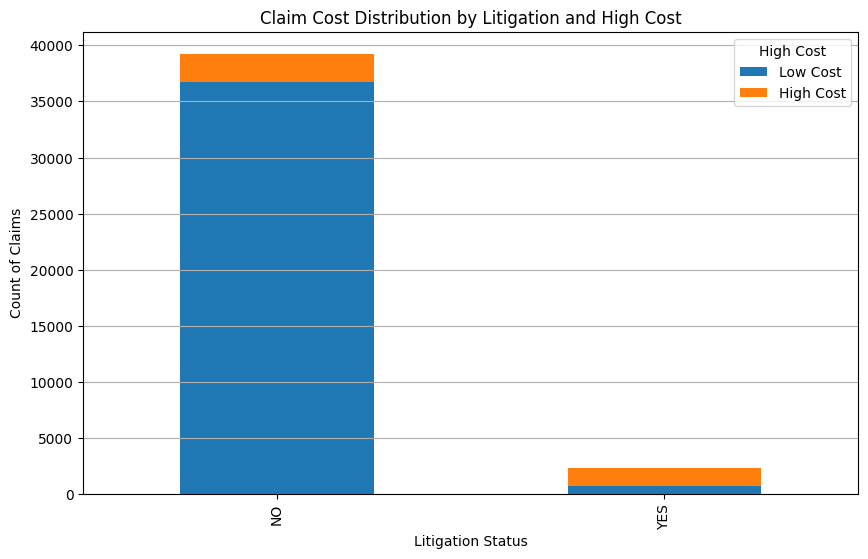
****

- Significant gap in claim costs between litigation and non-litigation for certain occupations, such as **pilot, utility, tennis, engineering, and room attendant, warrants a deeper investigation by the insurer**

- Understanding whether these gaps are due to particular risks or hazards associated with these roles could provide insights

- Consider if there are policy adjustments that could be made to address the higher litigation costs. This might include clearer policy terms, improved communication with policyholders, or additional coverage options.

1. **Claim Count vs [High Cost and Litigation]**

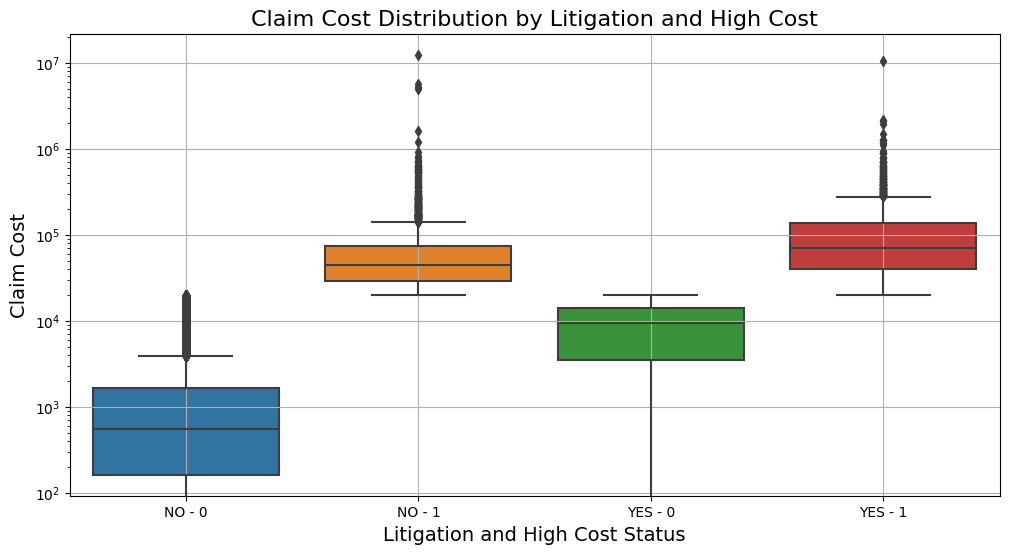


**High Cost – No Litigation records : 2528**

**High Cost – Litigation records : 1626**

**- Number of claims with no litigation but high cost are more than number of claims with litigation and high cost.**

1. **Claim Count vs [High Cost and Litigation]**

****

**Interpretation:**

- The distribution of claim costs for litigated claims (both YES - 0 and YES - 1) generally shows higher values compared to non-litigated claims.

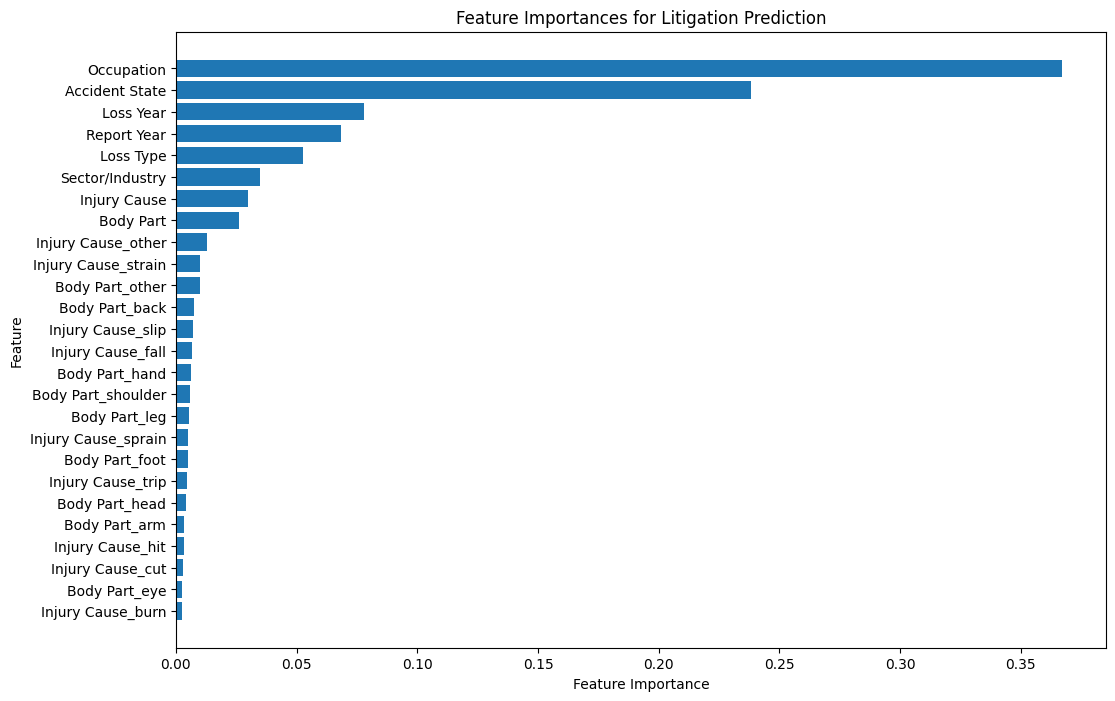
- The median and the interquartile range (IQR) of high-cost claims without litigation (NO - 1) are notably higher than the low-cost claims without litigation (NO - 0).

**Insights:**

- **Litigation tends to increase claim costs, especially when the claim is already categorized as high cost.**

- Non-litigated high-cost claims are also significant, indicating that **high costs are not exclusively driven by litigation.**

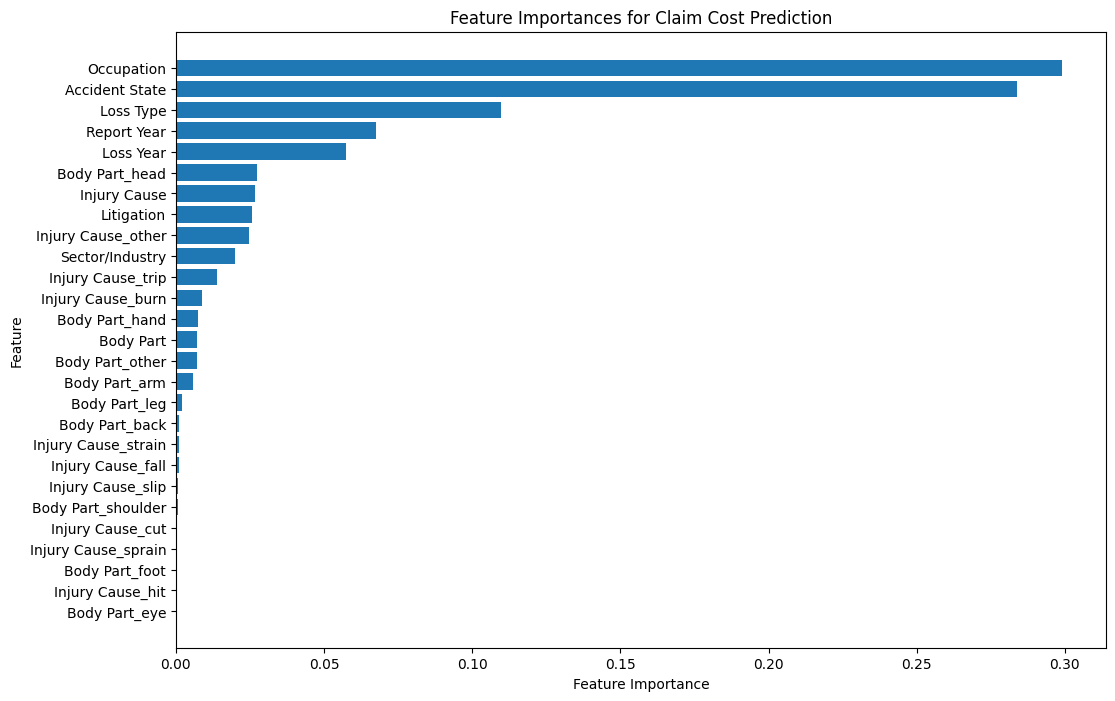
**Most Important Features for Litigation – Yes or No**

****

**Namely in descending order of importance :**

1. **Occupation**
2. **Accident State**
3. **Loss type**
4. **Sector/Industry**
5. **Injury Cause**
6. **Body Part**

**Most Important Features for Claim Cost Prediction**

****

**Namely in descending order of importance :**

1. **Occupation**
2. **Accident State**
3. **Loss type**
4. **Sector/Industry**
5. **Injury Cause**