# CSI Business Analyst Assessment

- From Siri Kademani

# Contents

## ➢ Data Cleaning

**Performed following steps in data preparation:**

1. Loaded data into Power BI via given CSV file.

2. Converted ‘Stories’ Column into ‘Text’ data type in Power query as there were some rows which were not getting loaded due to data type mismatch in Stories column. Loaded the data after changing the data type.

3. Checked if all columns are linked to the right data type and converted those which weren’t having right data type to the correct data type.

**Performed following operations to simplify dashboard and analysis.**

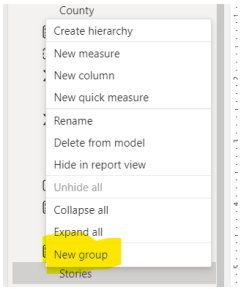
4. Created a new column ‘TIV’ by calculating the Total Insured Value [TIV] using the given formula.

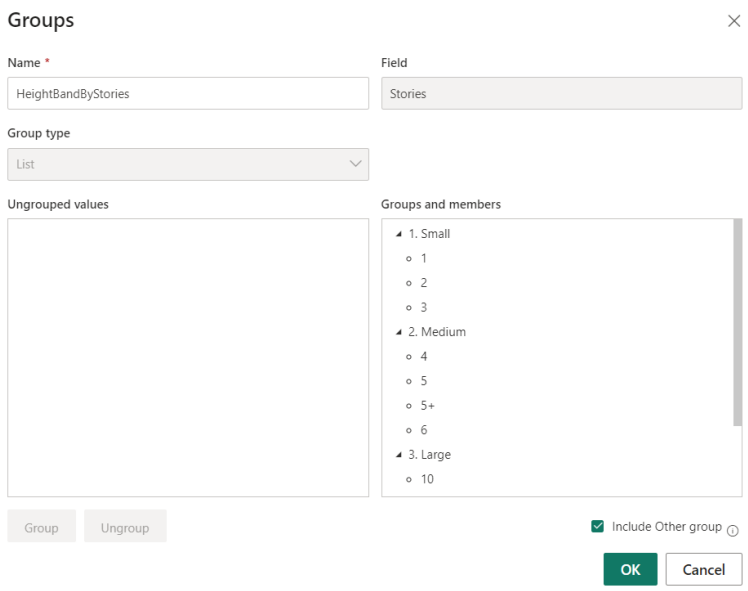
```
TIV = WDI_Analytical_Assessment_Data[Building Value]+WDI_Analytical_Assessment_Data[Other Value]+WDI_Analytical_Assessment_Data[Contents Value]
+WDI_Analytical_Assessment_Data[Time Element Value]
```

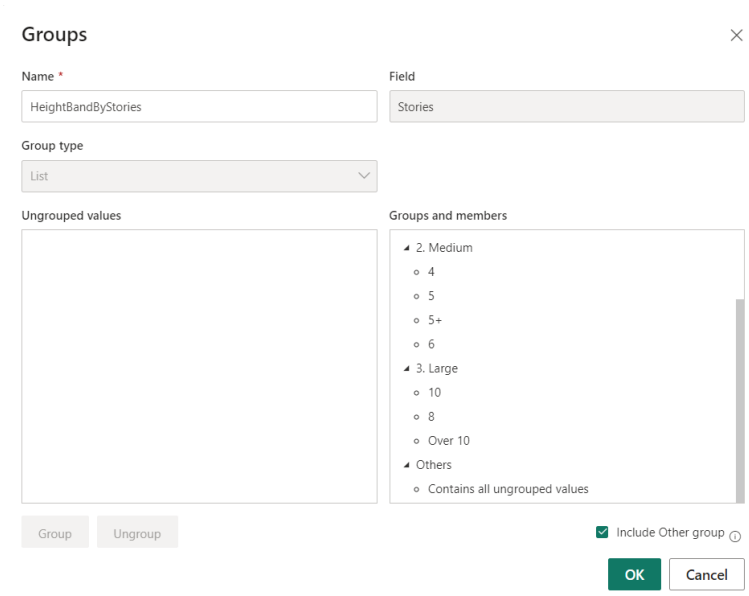
5. Created a new column ‘Risk Count’ with a default value of 1.

```
RiskCount = 1
```

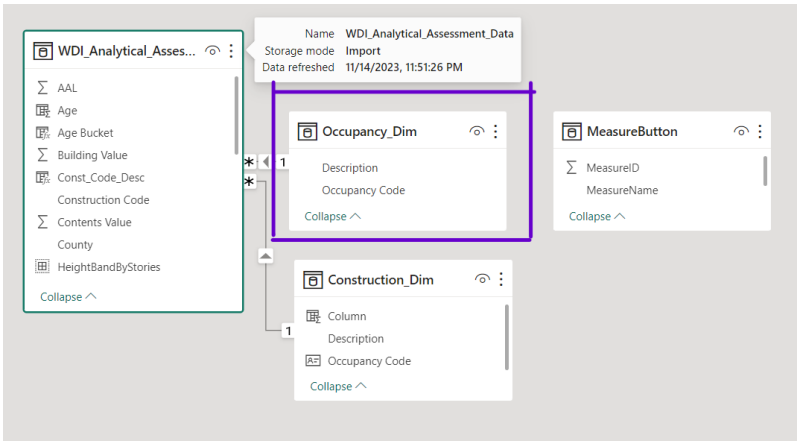
6. Created a new column ‘HeightBandByStories’ to categorize the ‘Stories’ column into ‘Small’, ‘Medium’ and ‘Large’ buckets. The number of stories within each height range was already given in the assessment overview doc.



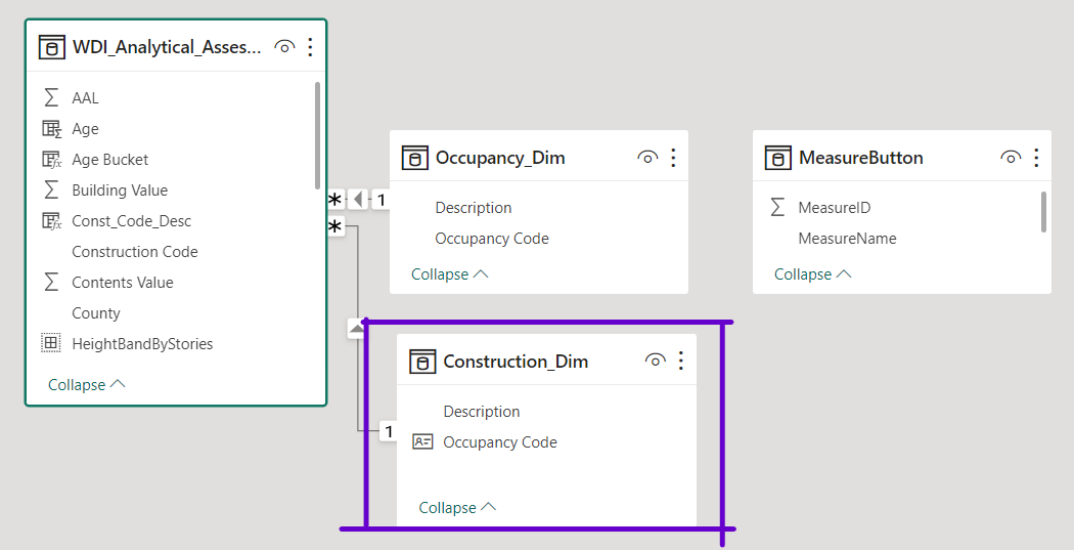



7. Created ‘Occupancy_Dim’ Table as a lookup table that contained ‘Occupancy_Code’ and ‘Description’ columns and linked it with the main ‘WDI_Analytical_Assessment_Data’ table. This was done to get description of Occupancy code. Adding Description to dashboard will help users to better understand the graphs.




| Occupancy Code | Description |
|---|---|
| ATC-00 | Unknown |
| ATC-01 | Single Family Housing |
| ATC-02 | Multi-Family Housing |
| ATC-37 | General Commercial |
| ATC-38 | General Industrial |

8. Created ‘Construction_Dim’ Table as a dimension table with columns ‘’Occupancy Code’ and ‘Description’ and linked it with the main table.
9. Renamed ‘Wood-Wood Frame’ to ‘Wood Frame’ for the code ‘WD10’ as the main table had value ‘Wood Frame’ for Construction_Code column. This was done to keep the data consistent.



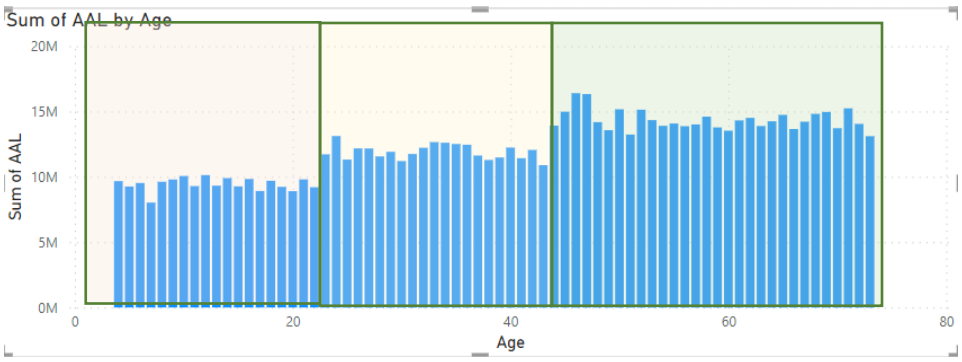

| Occupancy Code | Description |
|---|---|
| WD00 | Wood |
| WD10 | Wood Frame |
| MS00 | Masonry |
| MS10 | Masonry – Unreinforced |
| CN00 | Concrete |
| CN10 | Concrete – Reinforced |
| MH00 | Mobile Home |
| MH10 | Mobile Home – No Tie-Down |

10. Created ‘Age’ column used Today() and Year built column

```
1  Age = IF(NOT(ISBLANK(WDI_Analytical_Assessment_Data[Year Built])), YEAR(TODAY()) -WDI_Analytical_Assessment_Data[Year Built], BLANK())
```

| Stories | Year Built | AAL | TIV | RiskCount | TIV_WD10 | HeightBandByStories | Const_Code_Desc | Age | Age Bucket | TIV |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2000 | 9,987 | 594,467 | 1 | 0 | 1. Small | Mobile Home – No Tie-Down | 23 | 2. 23 to 44 yrs | |

11. Created ‘Age Bucket’ column by identifying the change in pattern of AAL values by age.




```
1  Age Bucket = IF(WDI_Analytical_Assessment_Data[Age] < 23, "1. Under 23 yrs", IF( WDI_Analytical_Assessment_Data[Age] < 45, "2. 23 to 44 yrs", "3. 45 yrs and above"))
```

| Stories | Year Built | AAL | TIV | RiskCount | TIV_WD10 | HeightBandByStories | Const_Code_Desc | Age | Age Bucket | TIV_Portfolio1 | TIV_Portfolio |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2000 | 9,987 | 594,467 | 1 | 0 | 1. Small | Mobile Home – No Tie-Down | 23 | 2. 23 to 44 yrs | 594467 | |

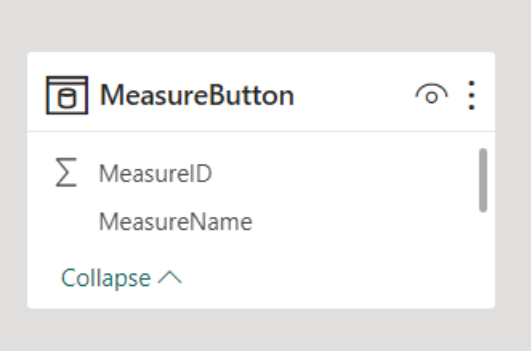
12. Created ‘TIV_WD10’ column for performing Task 2.

```
1  TIV_WD10 = IF(WDI_Analytical_Assessment_Data[Const_Code_Desc]="Wood Frame",WDI_Analytical_Assessment_Data[TIV],0)
```

| ccupancy Code | Construction Code | Stories | Year Built | AAL | TIV | RiskCount | TIV_WD10 | HeightBandByS |
|---|---|---|---|---|---|---|---|---|
| TC-01 | MH10 | 1 | 2000 | 9,987 | 594,467 | 1 | 0 | 1. Small |

13. Created Measure Button Column and ‘Dynamic Measure Column’ to prepare slicers for ‘Annual Average Loss’, Total Insured Value’ and ‘Risk Count’ in ‘Overview’ sheet. We have switch the measures in the graphs easily using this slicer.

```
1  Dynamic Measure = SWITCH(MeasureButton[Selected Measure],1,SUM(WDI_Analytical_Assessment_Data[AAL]),2,SUM
   (WDI_Analytical_Assessment_Data[RiskCount]),3,SUM(WDI_Analytical_Assessment_Data[TIV]))
```

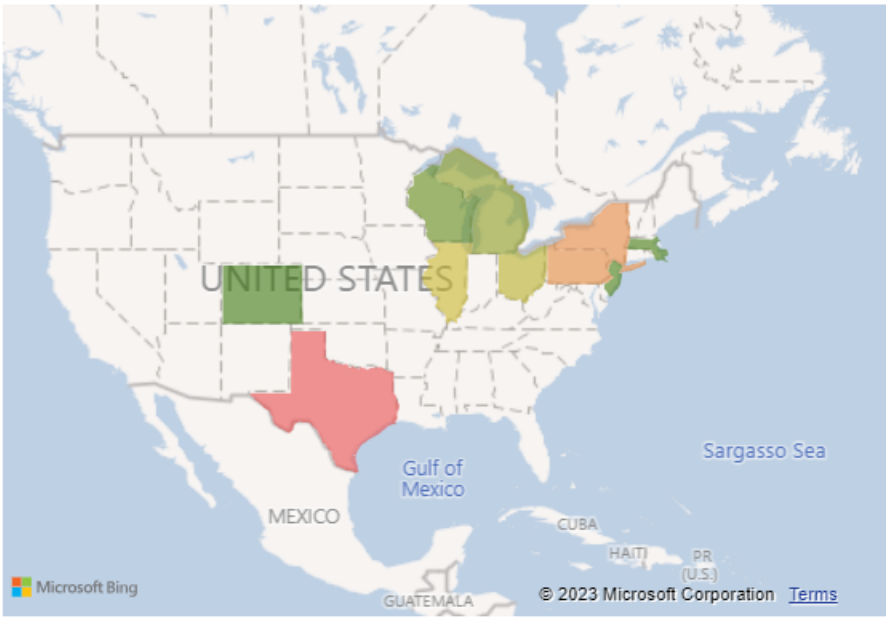


## ➢ Tasks

- Task 1

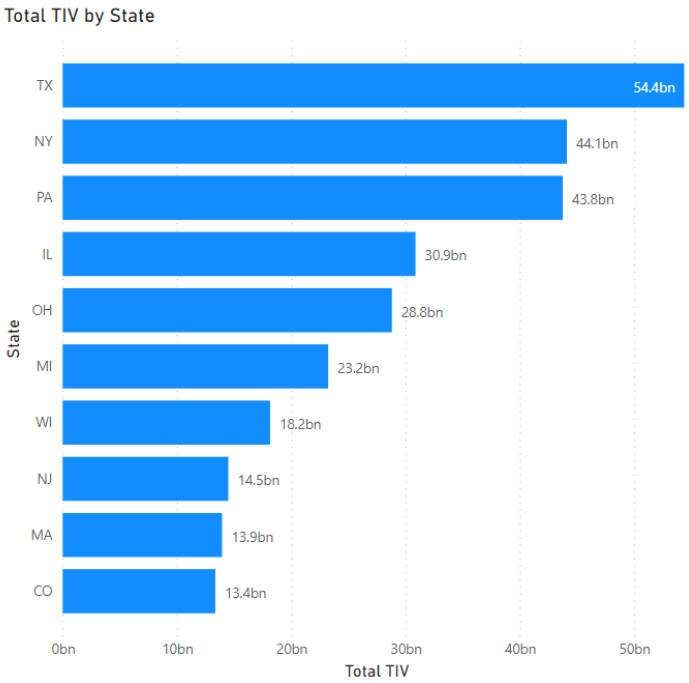
**For each state in the output file, find the total insured value (TIV) and number of risks.**

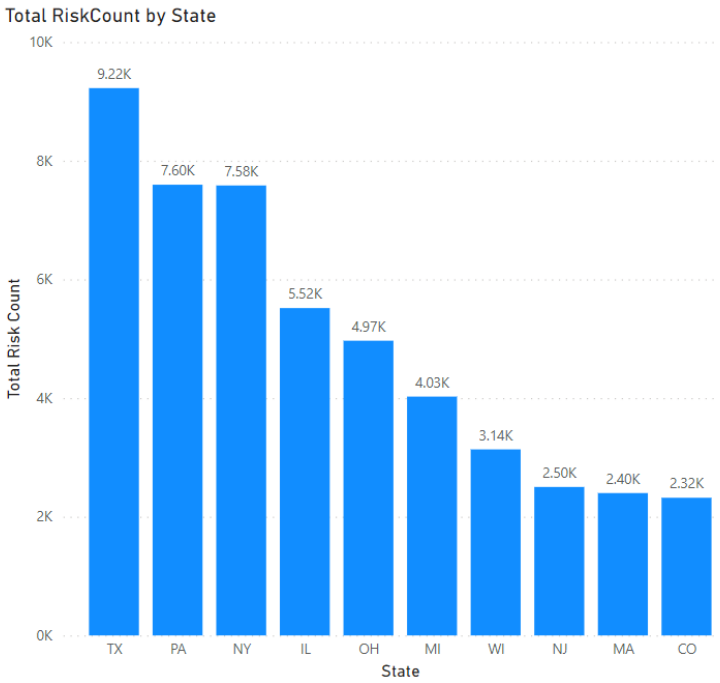
Average Annual Loss | Risk Count | **Total Insured Value**




Breakdown of AAL, Risk Count and TIV

| State | Total AAL | Average AAL | RiskCount | TIV |
|---|---|---|---|---|
| TX | 185M | 20,044 | 9,224 | 54,389M |
| PA | 127M | 16,668 | 7,596 | 43,767M |
| NY | 114M | 14,976 | 7,583 | 44,126M |
| IL | 100M | 18,041 | 5,517 | 30,877M |
| OH | 82M | 16,538 | 4,967 | 28,810M |
| MI | 73M | 18,016 | 4,025 | 23,235M |
| WI | 56M | 17,978 | 3,135 | 18,153M |
| CO | 46M | 19,784 | 2,320 | 13,356M |
| NJ | 43M | 16,996 | 2,502 | 14,496M |
| **Total** | **860M** | **17,461** | **49,268** | **285,156M** |



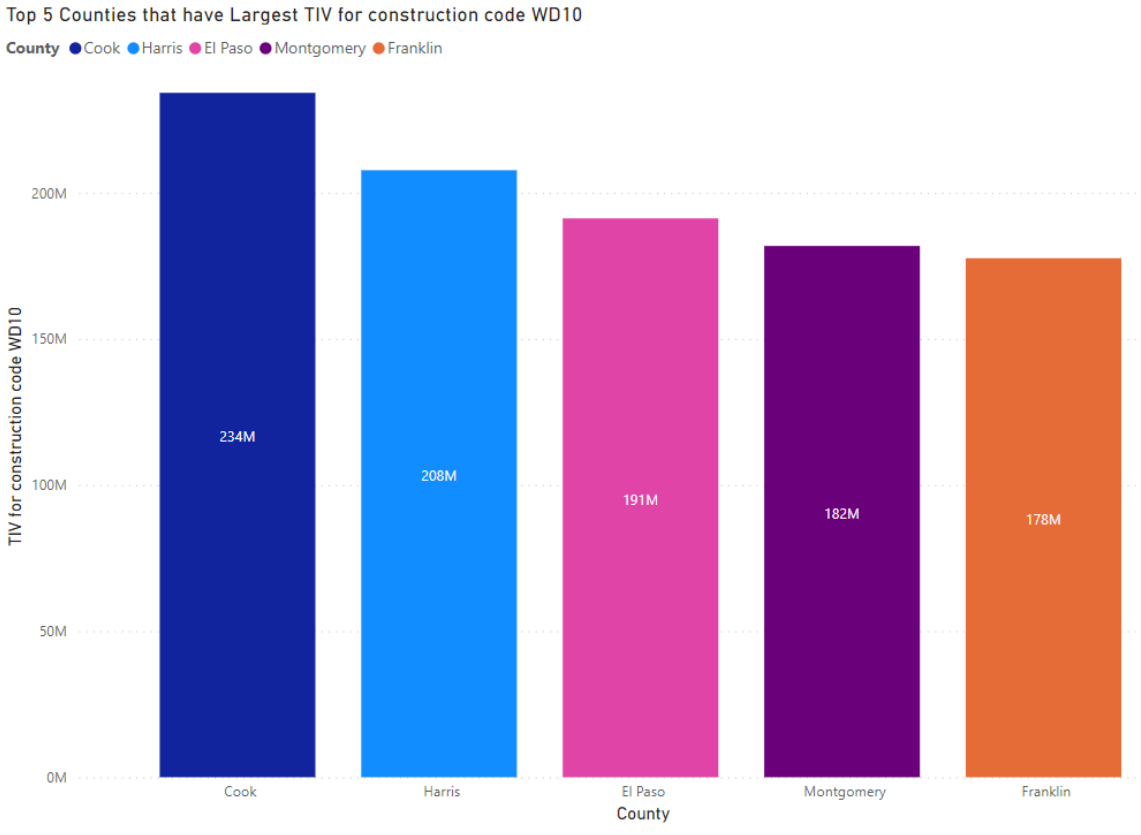




**TIV by state:** Texas has the highest total insured value while Colorado has the lowest.

**Risk Count by Sate:** Total Risk count for Texas is higher than other states while Colorado has the lowest risk count.

- Task 2

**Which 5 counties contain the largest total insured value (TIV) for construction code WD10?**



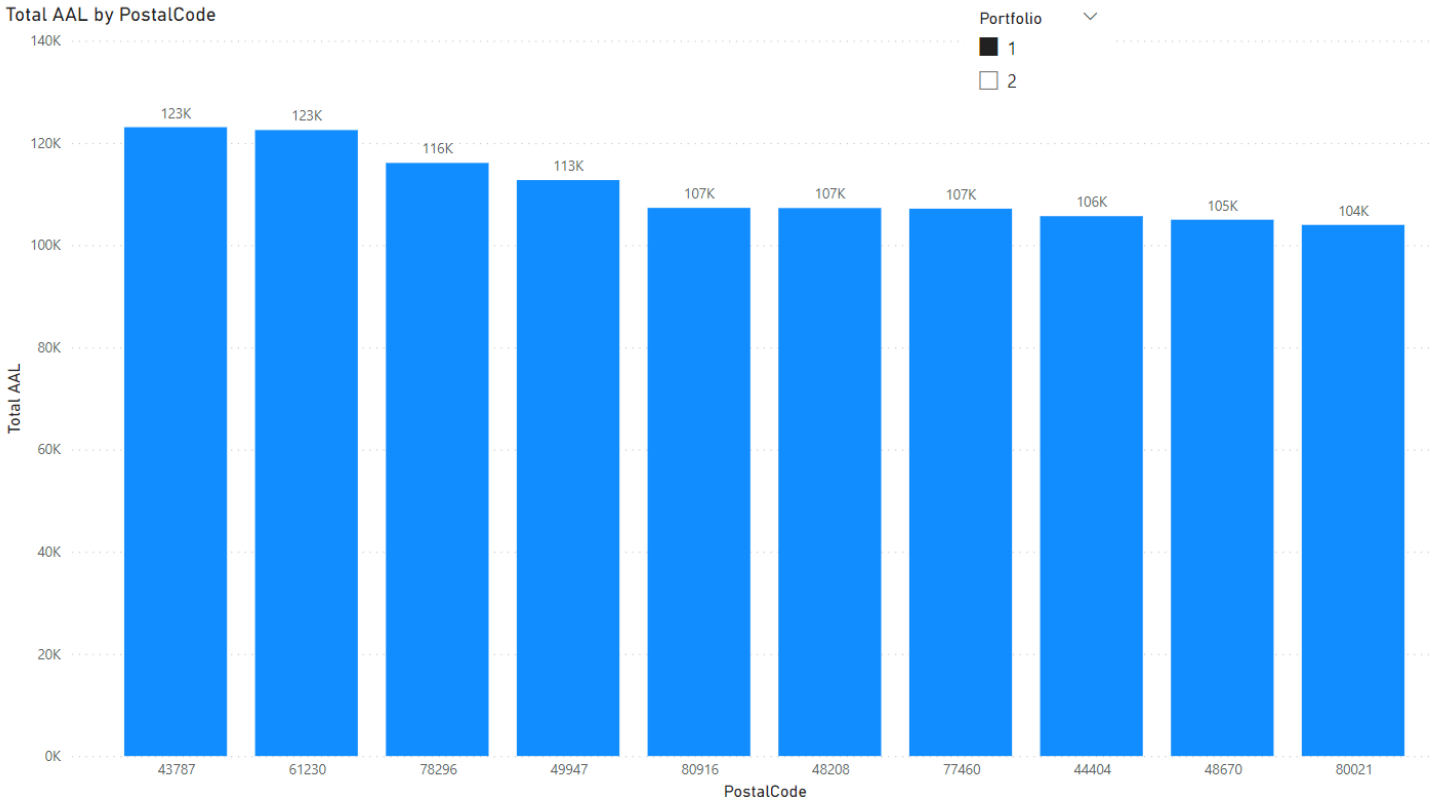
| County | TIV for construction code WD10 |
|---|---|
| Cook | 234,237,580 |
| Harris | 207,748,469 |
| El Paso | 191,267,328 |
| Montgomery | 181,825,426 |
| Franklin | 177,645,012 |

Cook, Harris, El Paso, Montgomery, Franklin are the 5 counties that have the largest Total Insured Values for 'WD10'.

- Task 3

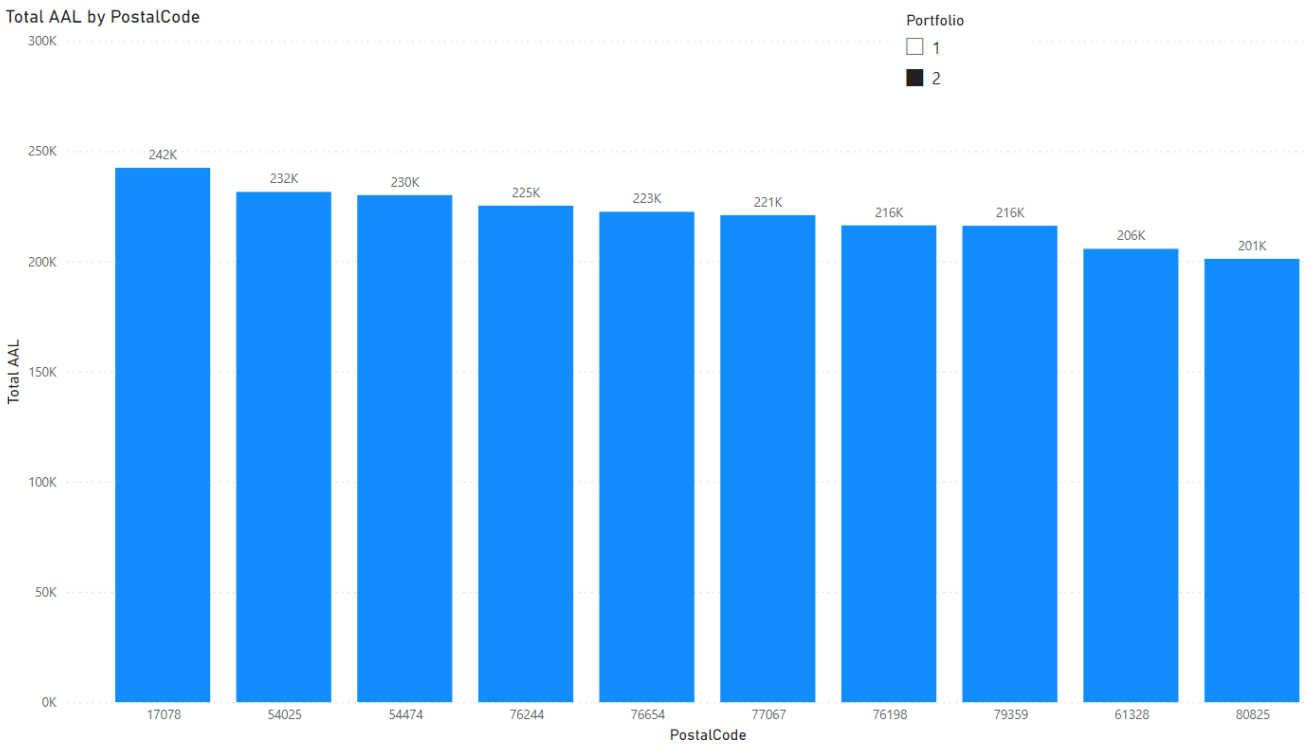
**For each portfolio, find the 10 postal codes that are the most susceptible to damage from winter storms.**

Top 10 postal codes for portfolio 1 that are most susceptible to damage.




| PostalCode | Total AAL |
|---|---|
| 43787 | 123,087 |
| 61230 | 122,526 |
| 78296 | 116,088 |
| 49947 | 112,737 |
| 80916 | 107,296 |
| 48208 | 107,260 |
| 77460 | 107,138 |
| 44404 | 105,673 |
| 48670 | 104,979 |
| 80021 | 103,967 |

Top 10 postal codes for portfolio 2 that are most susceptible to damage.




| PostalCode | Total AAL |
|---|---|
| 17078 | 242,459 |
| 54025 | 231,526 |
| 54474 | 230,070 |
| 76244 | 225,269 |
| 76654 | 222,514 |
| 77067 | 220,991 |
| 76198 | 216,351 |
| 79359 | 216,181 |
| 61328 | 205,704 |
| 80825 | 201,163 |

- Task 4

**For each Pennsylvania postal code, find the breakdown of total insured value (TIV), risk count, and average annual loss (AAL) by building height band.**

| PostalCode | Total Risk Count | Total of TIV | Total AAL | Average of AAL |
|---|---|---|---|---|
| **17078** | **9** | **102,272,299** | **253,160** | **28,129** |
| 1. Small | 7 | 70,876,495 | 158,510 | 22,644 |
| 3. Large | 1 | 19,644,021 | 56,575 | 56,575 |
| 2. Medium | 1 | 11,751,783 | 38,076 | 38,076 |
| **15475** | **9** | **101,546,233** | **233,043** | **25,894** |
| 1. Small | 5 | 51,837,840 | 173,651 | 34,730 |
| 3. Large | 2 | 20,936,126 | 33,498 | 16,749 |
| 2. Medium | 2 | 28,772,267 | 25,895 | 12,948 |
| **16853** | **11** | **64,260,273** | **229,725** | **20,884** |
| 1. Small | 7 | 38,203,506 | 139,286 | 19,898 |
| 2. Medium | 3 | 10,260,243 | 60,110 | 20,037 |
| 3. Large | 1 | 15,796,524 | 30,329 | 30,329 |
| **15951** | **10** | **82,047,302** | **222,931** | **22,293** |
| 3. Large | 2 | 38,861,389 | 86,648 | 43,324 |
| 1. Small | 5 | 16,986,158 | 74,248 | 14,850 |
| 2. Medium | 3 | 26,199,755 | 62,036 | 20,679 |
| **19104** | **10** | **63,995,459** | **220,287** | **22,029** |
| **18431** | **9** | **56,041,048** | **204,477** | **22,720** |
| **15565** | **5** | **69,943,148** | **200,854** | **40,171** |
| **17043** | **9** | **69,280,627** | **186,887** | **20,765** |
| **15047** | **5** | **58,341,118** | **183,751** | **36,750** |
| **17519** | **8** | **59,620,662** | **180,222** | **22,528** |
| **18218** | **7** | **60,274,572** | **179,006** | **25,572** |
| **15650** | **7** | **68,980,467** | **177,559** | **25,366** |
| **17579** | **6** | **77,610,348** | **177,261** | **29,544** |
| **18058** | **7** | **61,065,192** | **176,355** | **25,194** |
| **17086** | **7** | **69,599,654** | **173,239** | **24,748** |
| **17222** | **8** | **71,890,761** | **172,708** | **21,589** |
| **Total** | **7,596** | **43,766,679,186** | **126,609,013** | **16,668** |

State

PA

## ➢ Question 1.

**Does one of the portfolios appear to be more vulnerable to damage from winter storm? If so, which one? Explain your reasoning and any analysis you conducted to support your conclusion**.

**Answer**: Yes, upon thorough analysis, it is evident that Portfolio 2 is more susceptible to damage from winter storms compared to Portfolio 1.

- The key contributing factor to this vulnerability is the substantial difference in the estimated losses (AAL) between the two portfolios. Portfolio 2 exhibits a notably higher AAL, approximately **557** million, compared to Portfolio 1, which has an AAL of around **303** million.

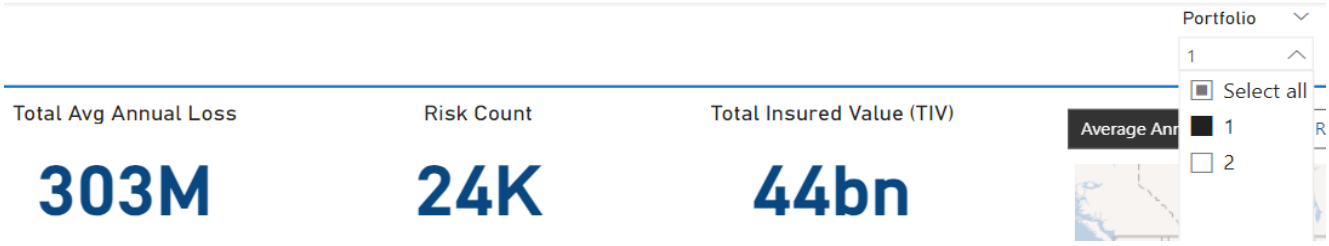



**Fig 1: KPIs for Portfolio 1**

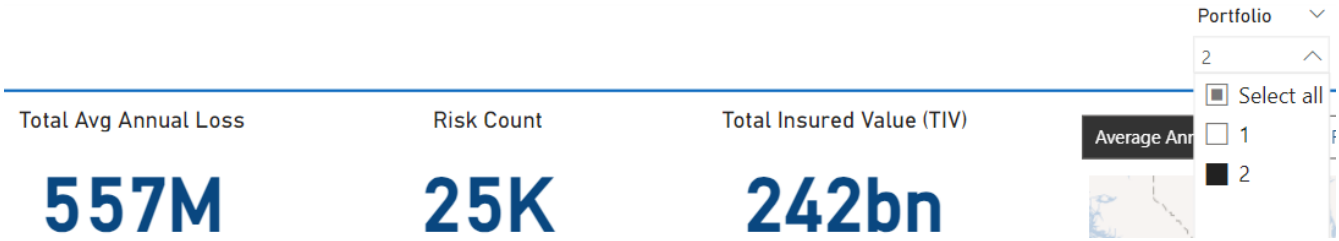



**Fig 2: KPIs for Portfolio 2**

- Despite both portfolios covering a similar number of risk counts (Risk count for Portfolio 1: 24K; Risk count for Portfolio 2: 25K), the discrepancy in the average of annualized average estimated losses, particularly a difference of 10,000, underscores the heightened vulnerability of Portfolio 2.

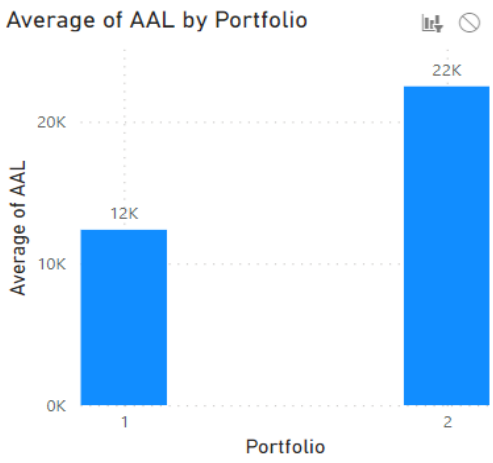



**Fig: Mean value of AAL for both Portfolios**

- Additionally, it's crucial to note that the standard deviation for AAL in Portfolio 1 is 6.3 thousand, indicating a moderate level of variability in the losses. And the higher standard deviation of 13.4 thousand for AAL in Portfolio 2 highlights a ***more significant level of risk in losses***.

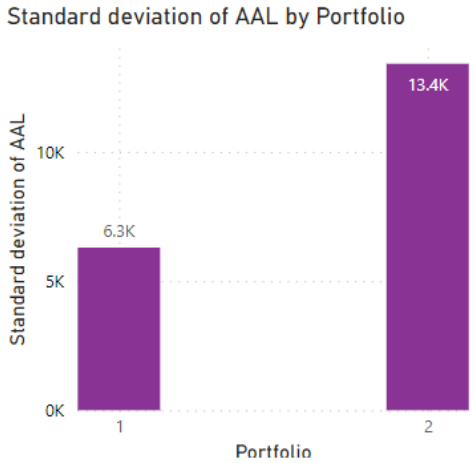



**Fig: Standard Deviation of AAL for both Portfolios**

We observed that ***TIV is positively correlated with AAL***. On analyzing TIV and AAL distributions for both the portfolios, we observed that:

- The TIV distribution in Portfolio 1 is concentrated below 1.5 million, indicating a prevalence of lower-valued insured properties with potentially lower replacement costs. In contrast, Portfolio 2 demonstrates a distinct pattern, with TIV values primarily falling within the higher range of 2.5 million to 3.5 million, suggesting a prevalence of higher-valued assets in this portfolio.

- Examining AAL ranges, Portfolio 1 demonstrates predominantly lower values (5,000 to 6,000), indicating generally less severe losses per risk. In contrast, ***Portfolio 2 has higher AAL values (8,000 to 10,000), suggesting potentially more significant and costly losses compared to Portfolio 1***.

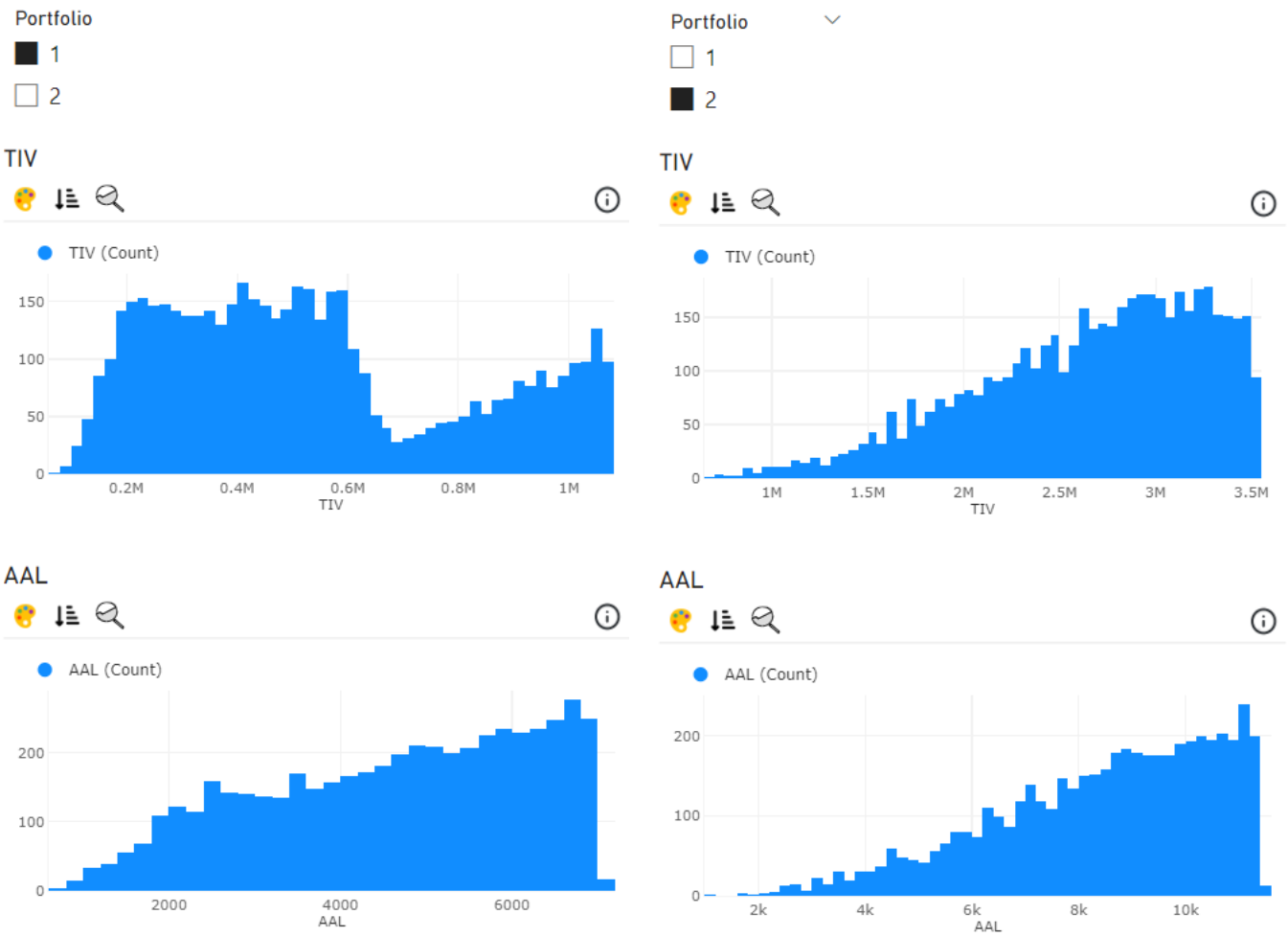



**Fig: TIV and AAL Distribution for Portfolio 1 and 2.**

The TIV distribution and AAL patterns reveal distinct risk profiles, with Portfolio 1 characterized by lower-valued insured properties and less severe losses, while Portfolio 2 exhibits higher-valued assets and potentially more significant and costly losses.

These nuanced insights, including the variability in AAL, further support the conclusion that Portfolio 2 is more exposed to potential damage from winter storms.

## ➢ Question 2.

**Which factors appear to have the greatest influence on AAL? Explain your reasoning and any analysis you conducted to support your conclusion.**

**Answer:**

On analyzing the data, we see that **TIV has the greatest influence over AAL**. TIV is positively correlated with AAL i.e., Higher TIV is associated with a higher AAL.

Other factors that have **influence on AAL are occupancy type, and construction materials**.

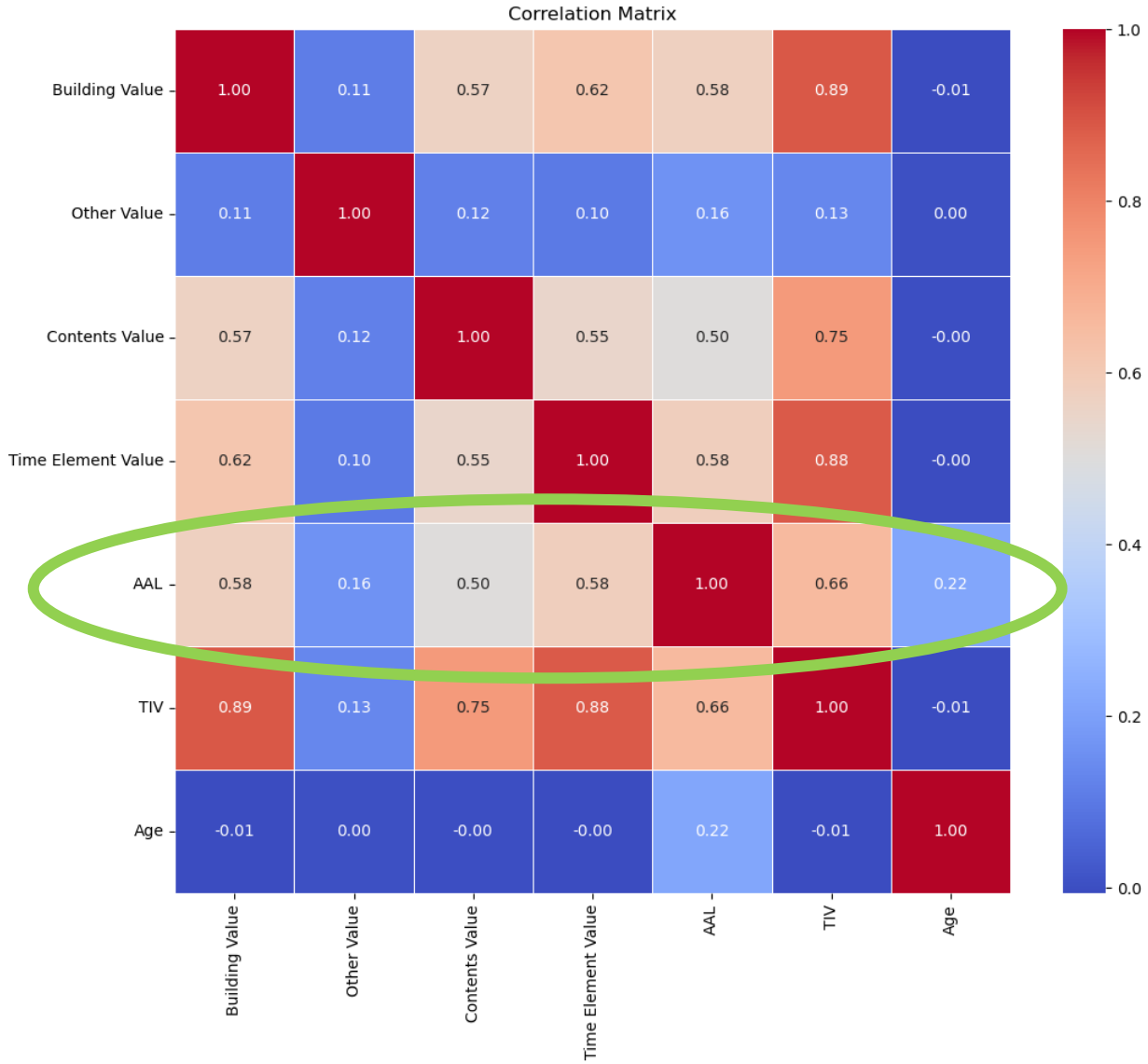



**Fig: Correlation matrix between AAL and other columns [Fig from Python file attached in the mail]**

**Factors which are influencing AAL are as follows:**

- **TIV**:
  - We can observe that large commercial or industrial property has a higher TIV than Residential property and thus the expected loss for Commercial properties is observed to be higher than residential property.

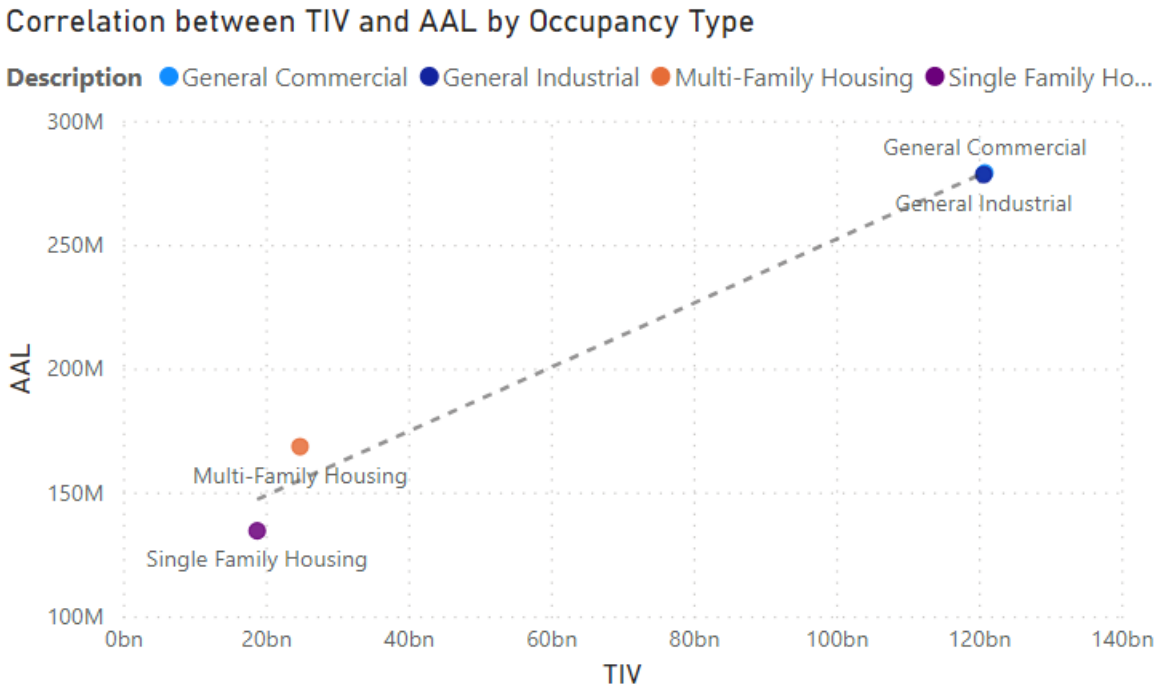



**Fig: Correlation between TIV and AAL**

- **Occupancy Type**:
  - The average annual loss for general commercial properties and for industrial properties are around 280M which are both on significantly higher end whereas the estimated losses are 168M and 134M for multi-family and single-family housing which are on lower end of the graph. Occupancy type plays a crucial role, as properties engaged in high-risk activities may require specialized construction, influencing both TIV and Average Annual Loss (AAL).

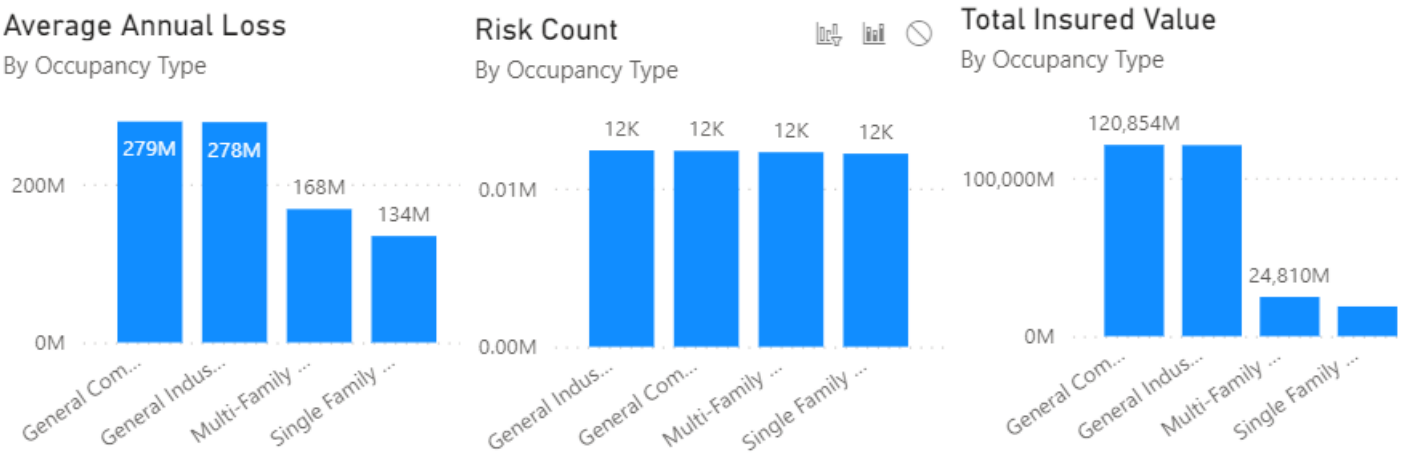



**Fig: Different KPIs by Occupancy Type**

➢ **Construction Materials**:

- Since properties with construction materials like wood, wood-frames, masonry, mobile homes – no tie down, mobile homes are easier and cheaper to build as compared to commercial/industrial properties which are built with concrete.

- High-value properties, often constructed with premium materials like concrete, command a higher Total Insured Value (TIV). In specialized constructions, such as commercial or industrial facilities, the use of concrete contributes to increased TIV due to elevated construction costs and unique design requirements. Thus, any damage to the property will cause a higher loss.

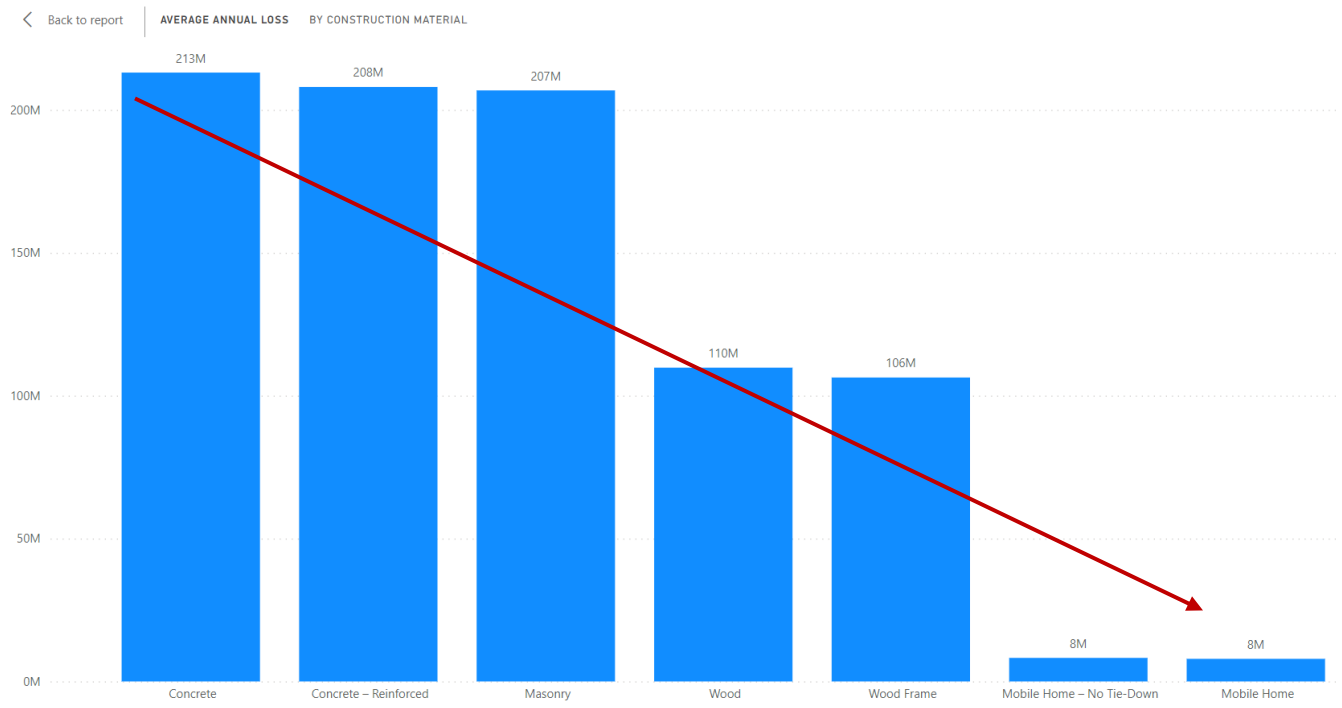
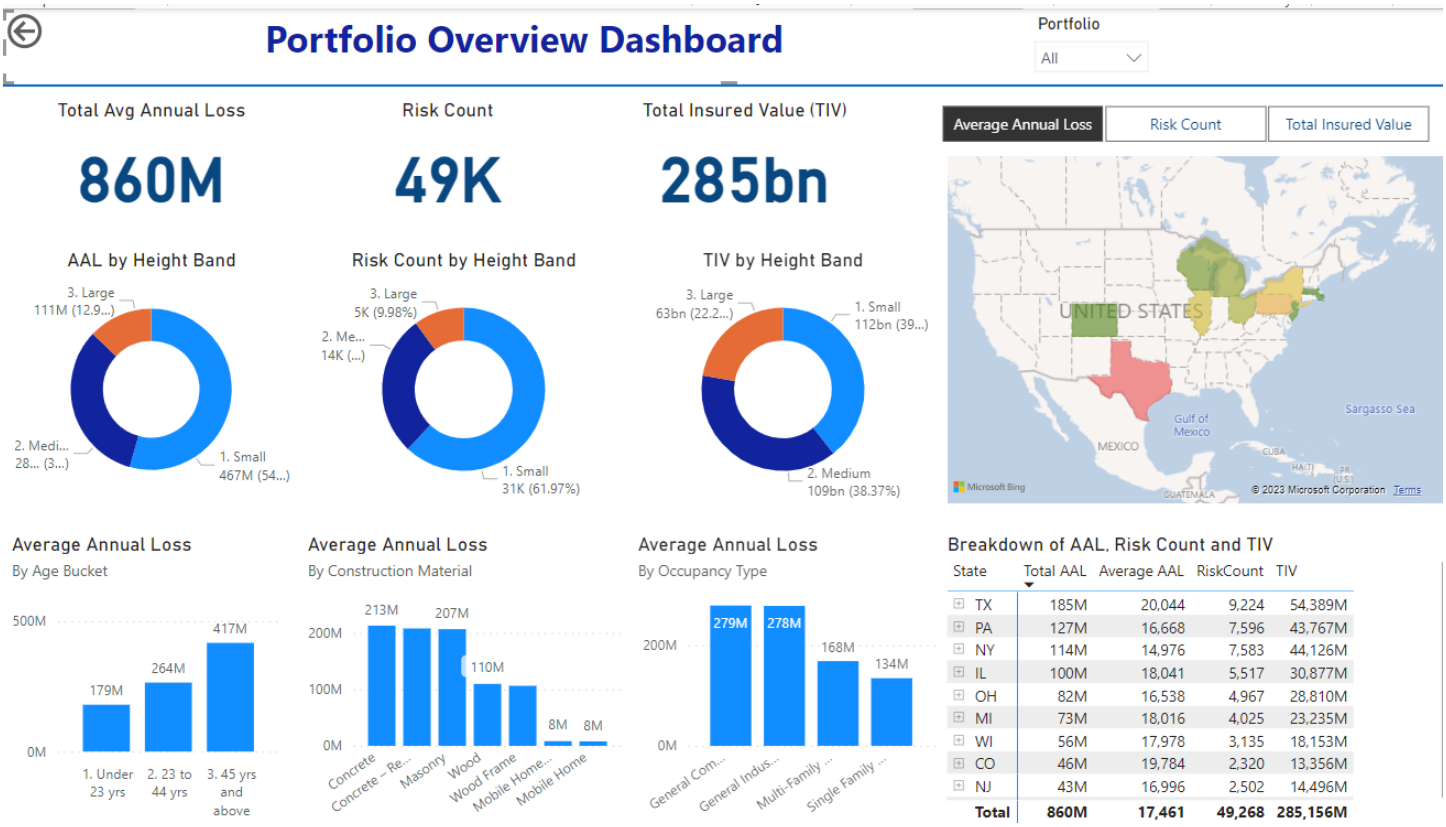



**Fig: AAL by construction materials.**

# ➢ Appendix

- Portfolio Overview Dashboard



- Portfolio 1 Overview

## Portfolio Overview Dashboard

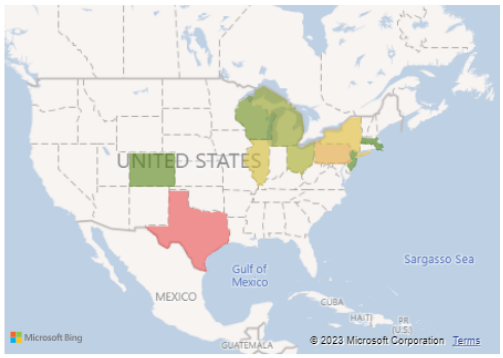
Portfolio: 1

Total Avg Annual Loss: **303M**

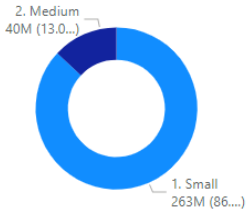
Risk Count: **24K**

Total Insured Value (TIV): **44bn**

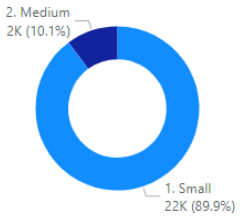
Average Annual Loss | Risk Count | Total Insured Value



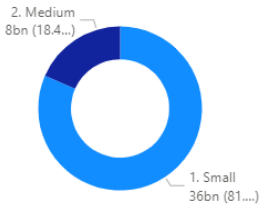

AAL by Height Band



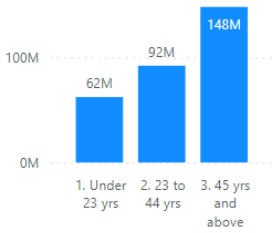

Risk Count by Height Band



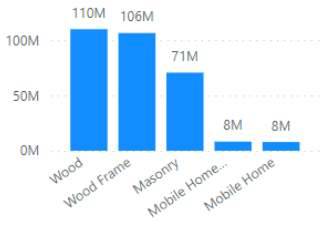

TIV by Height Band



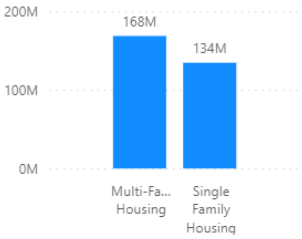

Average Annual Loss
By Age Bucket




Average Annual Loss
By Construction Material




Average Annual Loss
By Occupancy Type




Breakdown of AAL, Risk Count and TIV

| State | Total AAL | Average AAL | RiskCount | TIV |
|---|---|---|---|---|
| TX | 64M | 14,092 | 4,528 | 8,067M |
| PA | 45M | 11,785 | 3,797 | 6,823M |
| NY | 39M | 10,615 | 3,707 | 6,618M |
| IL | 36M | 12,946 | 2,758 | 4,880M |
| OH | 29M | 11,802 | 2,475 | 4,411M |
| MI | 26M | 12,872 | 2,024 | 3,613M |
| WI | 20M | 12,649 | 1,576 | 2,764M |
| CO | 17M | 14,227 | 1,169 | 2,086M |
| NJ | 15M | 11,813 | 1,261 | 2,254M |
| **Total** | **303M** | **12,373** | **24,484** | **43,610M** |

- Portfolio 2 Overview

## Portfolio Overview Dashboard

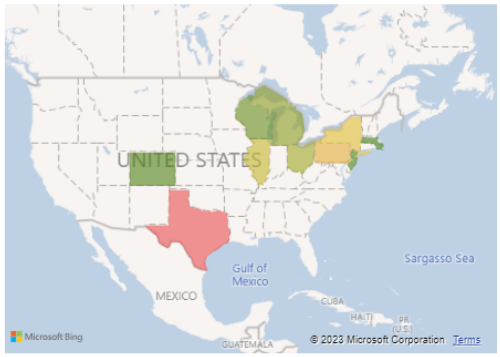
Portfolio: 2

Total Avg Annual Loss: **557M**

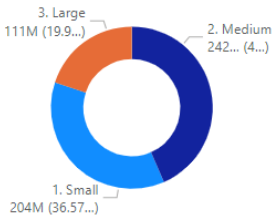
Risk Count: **25K**

Total Insured Value (TIV): **242bn**

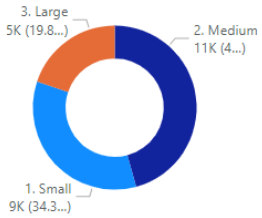
Average Annual Loss | Risk Count | Total Insured Value



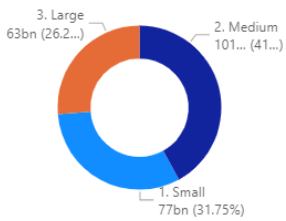

AAL by Height Band



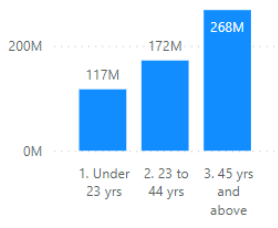

Risk Count by Height Band



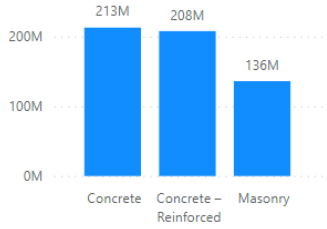

TIV by Height Band



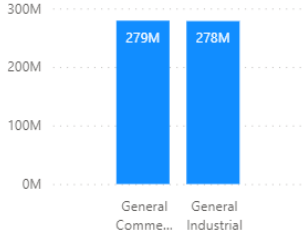

Average Annual Loss
By Age Bucket




Average Annual Loss
By Construction Material




Average Annual Loss
By Occupancy Type




Breakdown of AAL, Risk Count and TIV

| State | Total AAL | Average AAL | RiskCount | TIV |
|---|---|---|---|---|
| TX | 121M | 25,782 | 4,696 | 46,322M |
| PA | 82M | 21,548 | 3,799 | 36,944M |
| NY | 74M | 19,147 | 3,876 | 37,509M |
| IL | 64M | 23,135 | 2,759 | 25,997M |
| OH | 53M | 21,241 | 2,492 | 24,399M |
| MI | 46M | 23,220 | 2,001 | 19,622M |
| WI | 36M | 23,365 | 1,559 | 15,388M |
| CO | 29M | 25,428 | 1,151 | 11,271M |
| NJ | 28M | 22,262 | 1,241 | 12,241M |
| **Total** | **557M** | **22,488** | **24,784** | **241,546M** |