**First of all, I chose this "Airline Passenger Satisfaction" dataset because airline statistics intrigued me. Also I chose the "Flight Distance" column because I think that there are interesting statistical relations between that passenger satisfaction and flight distance. This column shows us the flight distance of all passengers. According to the column information, the mean is 119k, standard deviation is 997 and quantiles are Q1 = 414, Q2(Median) = 844, Q3 = 1744. Also, we see the minimum value is 31 and the maximum value is 4983.**

**Content:**

**1. Mean of Column Data**

Mean = ∑ Flight Distance / n = **1190**

**2. Median of Column Data**

The median is the middle value of the dataset when it is sorted. Since the median is **844**, this means half of the passengers had flight distances less than 844 and the other half had more.

**3. Variance, Standard Deviation and Standard Error**

**Variance(σ²):** Measures how spread out the values are

**Standard Deviation(σ):** Square root of variance, shows average deviation from the mean.

Given: σ = 997

**Standard Error(SE):**

SE = Standard Deviation / sqrt(n)

= 997 / sqrt(129880)

= 997 / 360.38

= 2.77

**4. Decide the Shape of Distribution**

To determine the shape:

* Mean = 1190
* Median = 844

Since mean > median, the distribution is Right-Skewed (Positive Skewed).

**5. Find Outliers(Using IQR Method)**

* Compute IQR:
  + IQR = Q3 - Q1 = 1744 - 414 = 1330
* Calculate fences:
  + Lower fence: Q1 - 1.5 \* IQR = 414 - 1.5 \* 1330 = 414 - 1995 = -1581
  + Upper fence: Q3 + 1.5 \* IQR = 1744 + 1.5 \* 1330 = 1744 + 1995 = 3739

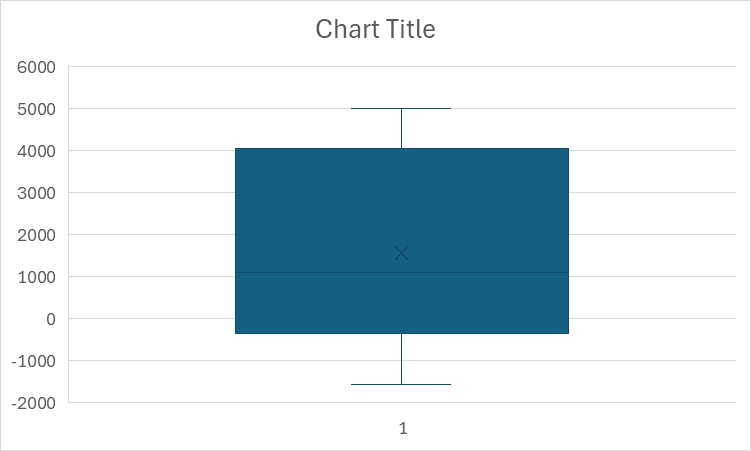
**Outliers:** Any values > 3739. From the data, since the maximum is 4983, there are outliers above the upper fence.

**6. Graph the Column Data and Comment**

To visualize the data, a histogram or density plot is ideal. You'd likely observe:

* Most values clustered between 200–2000.
* A long right tail due to some very large flight distances → confirming **Right-Skewness**.

**7.Boxplot**



A **boxplot** visually shows:

* Median
* IQR (box width)
* Minimum and maximum (excluding outliers)
* Outliers (as dots beyond whiskers)

The data is **widely spread**, ranging approximately from **-1500** to **5000**. The median(center line) is slightly above **1000**, suggesting a **Right-Skewed distribution** and the **presence** of **outliers** above **3739**.

**Final Conclusion:** The flight distance data shown:

* A Right-Skewed Distribution
* Presence of Outliers
* High Variability (Standard Deviation = 997)

This makes it an interesting variable to explore in relation to **passenger satisfaction**, especially to see if longer flights correlate with more (or less) satisfaction.