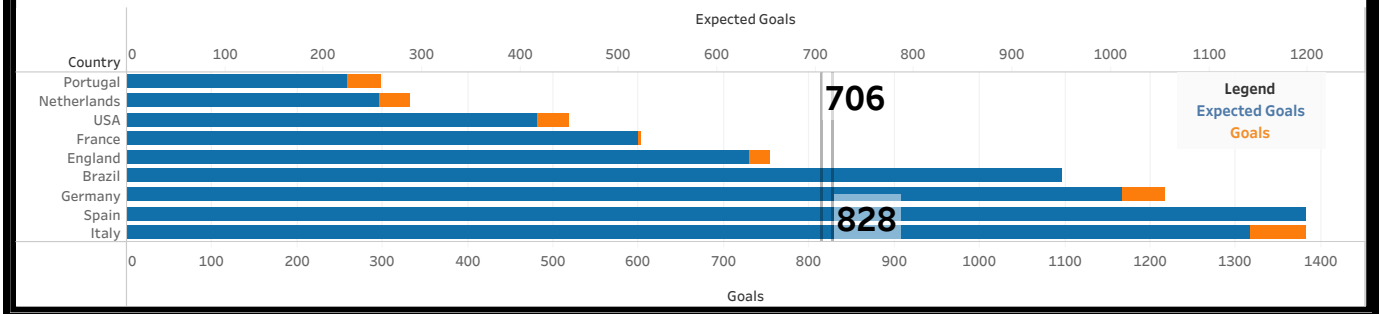
A statistical analysis of the efficacy of **expected goals** as a predictor ..

Expected Goals (xG) is one of the **most popular statistics** used in football. **Goals scored** is the most important statistic to anybody concerned with football. Expected goals therefore has become a very useful tool in football hence why I have decided to explore the **efficacy of this measure** as a predictor of **goals scored**.

How have **Expected Goals** v **Goals Scored** differed between 2016 - 2020 across the World?

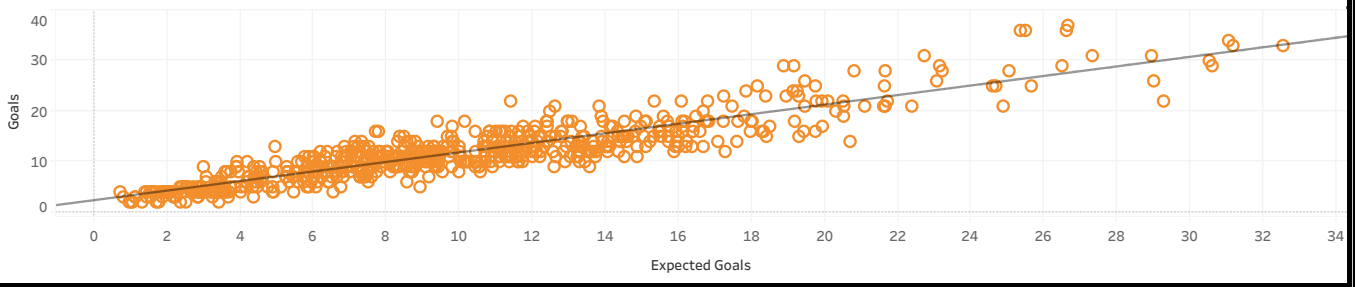


Between 2016 - 2020 the difference between the number of average **goals scored** and the number of average **expected goals** scored by professional football players in the **9 countries listed above** was **122 goals.** Given the context of professional football where goals scored are few, this is a notable finding.

What was the difference between **Expected goals** and **Goals scored** across the top 5 leagues in the world between 2016 - 2020?

The differences in **expected goals** and **goals scored** observed across the top five leagues in the world between **2016 - 2020** vary. The greatest difference of **241 goals** is shown in **Serie A in Italy** and the smallest difference of **4 goals** is shown in **France Ligue 1**. However **a key similarity** that I have observed between the five leagues is that **goals scored** exceeded **expected goals** across the **four year period** in **each of the five leagues.**

To what extent is **expected goals** a good predictor of **goals Scored**?



Key findings: Overall, between 2016 - 2020 **expected goals** exceeded **goals scored** in the 9 countries I have analysed apart from **Brazil.** However **goals scored** exceeded **expected goals** in the top five leagues in the world during the **same period**. Also, as per the **gradient** of the **plotted line** from this analysis **(0.95)**, **on average 2 expected goals was associated** with a profossenial footballer scoring **1.9 goals scored** in the 9 countries I have analysed between **2016 and 2020**.

Additionally, the **plotted line** above clearly shows a **strong positive relationship** between **expected goals** and **goals scored with few outliers**. Lastly, the p-value of **<0.0001** indicates that there is a less than **0.1% likelihood** that the relationship shown **along the plotted line** between **expected goals** and **goals s..**