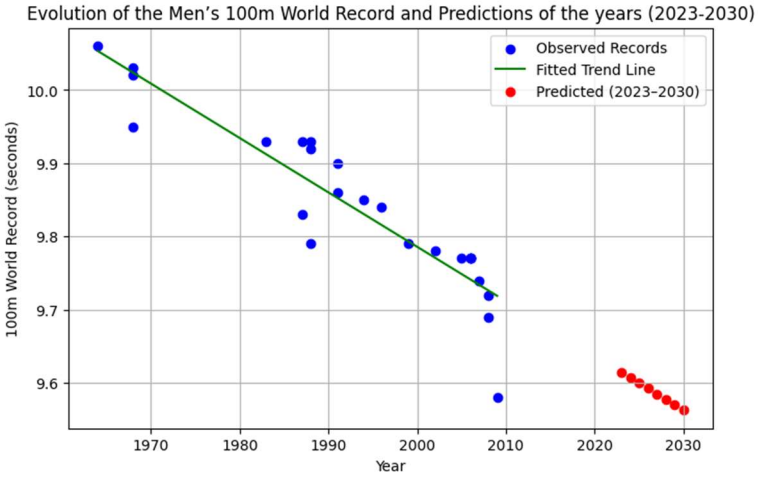
Interpretation of results

```
Intercept (b0):
 [24.64808957]
Slope (b1):
 [[-0.00743113]]
R² = 0.836, MSE = 0.00209
```



We see that the linear slope is decreasing at 0.00743. This means that every 4 years, the total time to run 100 m is decreasing by 0.03 s (or every year, it decreases by 7 ms). We also see how the impact of this slope is demonstrated in the predicted future world records. These results are good but should not be treated as true values, but as estimates based on performance, as it is impossible that the slope stays like this forever. Also, as time goes on, runners will plateau, and they won't be able to improve as much since it will not be physically possible. We see how the line goes across all the data points, but it is not very accurate, as some points are far from the trendline. We also have a statistical fit: R^2 is 0.836, which shows that it considers 84% of the variance in the data, which is good, but 16% is still missing. Although, the mean squared error is low at 0.00209, which means that it predicts values reasonably well.

```
Predicted 100m World Records:
  2023: 9.61 s
  2024: 9.61 s
  2025: 9.60 s
  2026: 9.59 s
  2027: 9.59 s
  2028: 9.58 s
  2029: 9.57 s
  2030: 9.56 s
```

Key Limitations

The first limitation that this model encounters is its linearity. This does not make sense whatsoever, since it is like saying that in a long time, someone will be able to run the 100 m in no time (or even negative time). A human usually plateaus with time, as it will not be physically possible to run 100 m in 2 seconds in the future. This is absurd, and I think another type of regression, like logistic regression, would have been more appropriate, as it would be adapted to when humans reach their own limit in speed development.

The regression also excludes wind assistance, the precision of equipment, and doping regulations. Including these factors might help build a more accurate model to estimate the proper slope. Air resistance is an important factor that was mentioned on the website but not accounted for in the regression. Also, if technology in shoes changes, that will have a direct impact on runners' times that is not covered. The same applies to the characteristics of the run (track type, weather, etc.).

We also have a small dataset, as we are only comparing the record-breaking years and not the average time of all professional runners per year. Having more values could lead to a better estimation if we observe how they improve yearly instead of only when there is a new record. This would create a larger sample size and yield more accurate results for the actual evolution of performance. It would also account for major outliers like Usain Bolt. As we can see, he broke the record at the end, which was much faster than the predictions—was he an exception, or was this the trend? We could answer this with more data and error bounds.