

FIFA Top 5% Player Performance Analysis (2017–2021)

This notebook explores how player attributes like **preferred foot**, **nationality**, **age**, **acceleration**, **agility**, and **BMI** impact **wage** and **potential rating** for the **top 5% of FIFA players** from 2017 to 2021.

Objective

To validate the hypothesis that:

"The top 5% of FIFA 21 players are faster (higher acceleration and agility) than those in FIFA 17."

Dataset Source: [Sofifa via Kaggle](#)

```
In [1]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')

# Load data from 'Data' folder
players17_df = pd.read_csv('Data/players_17.csv')
players18_df = pd.read_csv('Data/players_18.csv')
players19_df = pd.read_csv('Data/players_19.csv')
players20_df = pd.read_csv('Data/players_20.csv')
players21_df = pd.read_csv('Data/players_21.csv')

# Add season labels
players17_df["season"] = 2017
players18_df["season"] = 2018
players19_df["season"] = 2019
players20_df["season"] = 2020
players21_df["season"] = 2021

# Top 5% filter
def top_5_percent(df):
    return df.nlargest(int(0.05 * len(df)), 'overall')

dfs = [top_5_percent(df) for df in [players17_df, players18_df, players19_df, players20_df, players21_df]]
df = pd.concat(dfs)

# Calculate BMI
df['BMI'] = df['weight_kg'] / ((df['height_cm'] / 100) ** 2)
df.rename(columns={
    'movement_acceleration': 'acceleration',
    'movement_agility': 'agility',
    'short_name': 'name',
    'wage_eur': 'wage'
}, inplace=True)
df.drop_duplicates(inplace=True)
```

```
df.reset_index(drop=True, inplace=True)
df.head()
```

Out [1]:

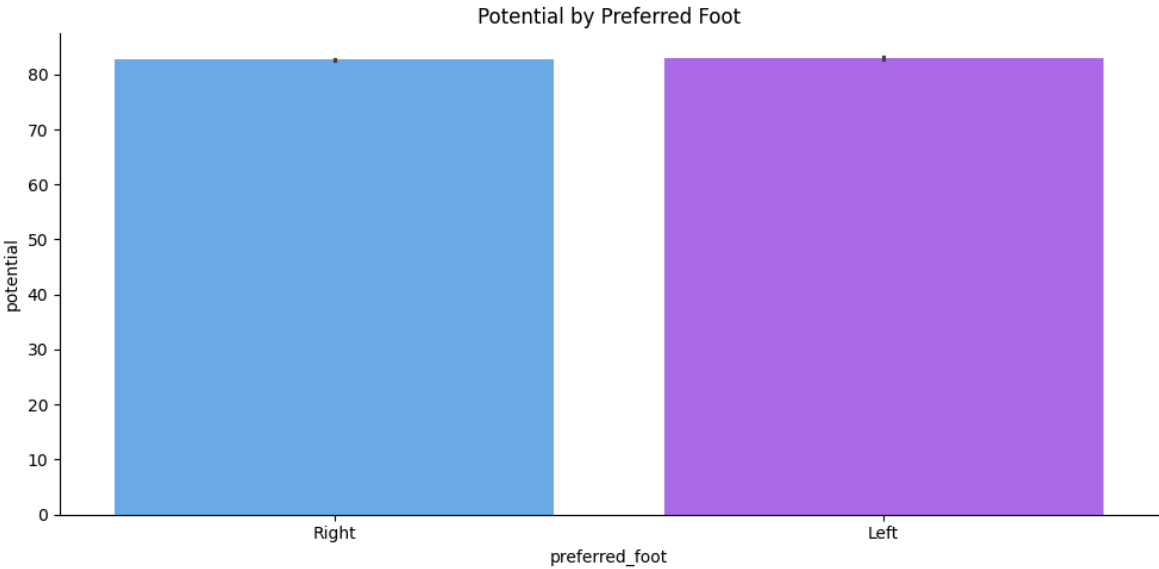
	sofifa_id	player_url	name	long_name	age	do
0	20801	https://sofifa.com/player/20801/cristiano-ronaldo-dos-santos-aveiro	Cristiano Ronaldo	Cristiano Ronaldo dos Santos Aveiro	31	1985-02-05
1	158023	https://sofifa.com/player/158023/lionel-messi-cuccittini	L. Messi	Lionel Andrés Messi Cuccittini	29	1987-06-26
2	190871	https://sofifa.com/player/190871/neymar-da-silva-santos-junior	Neymar	Neymar da Silva Santos Júnior	24	1992-02-15
3	167495	https://sofifa.com/player/167495/manuel-neuer	M. Neuer	Manuel Neuer	30	1986-03-19
4	176580	https://sofifa.com/player/176580/luis-suarez/1987-01-21	L. Suárez	Luis Alberto Suárez Díaz	29	1987-01-21

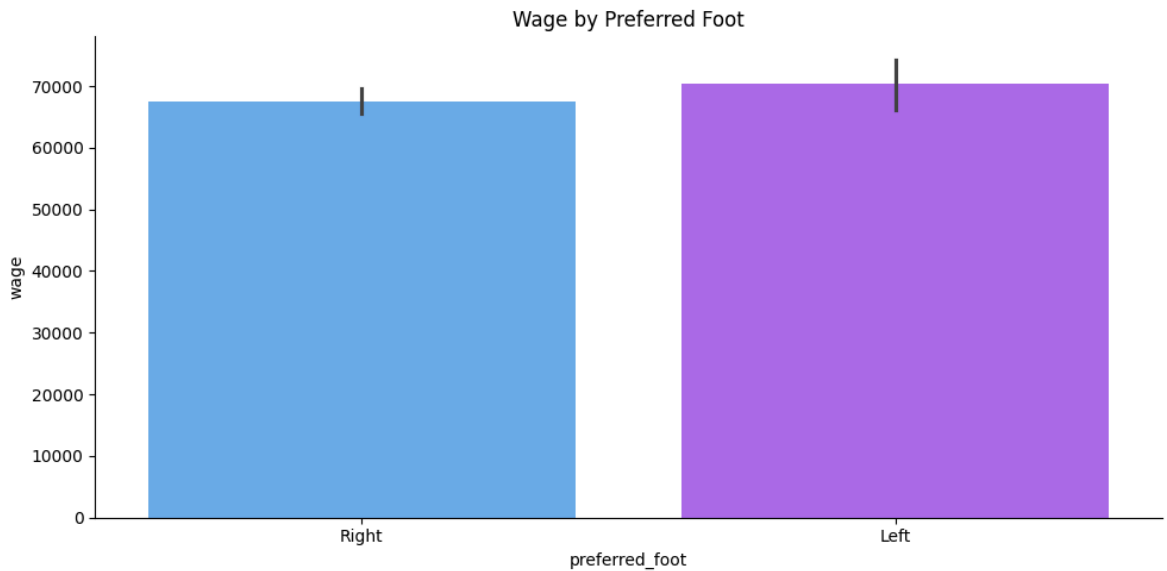
5 rows x 108 columns

Preferred Foot vs Potential and Wage

```
In [2]: sns.catplot(x='preferred_foot', y='potential', data=df, kind='bar', palette='c')
plt.title("Potential by Preferred Foot")
plt.tight_layout()
plt.show()

sns.catplot(x='preferred_foot', y='wage', data=df, kind='bar', palette='c')
plt.title("Wage by Preferred Foot")
plt.tight_layout()
plt.show()
```



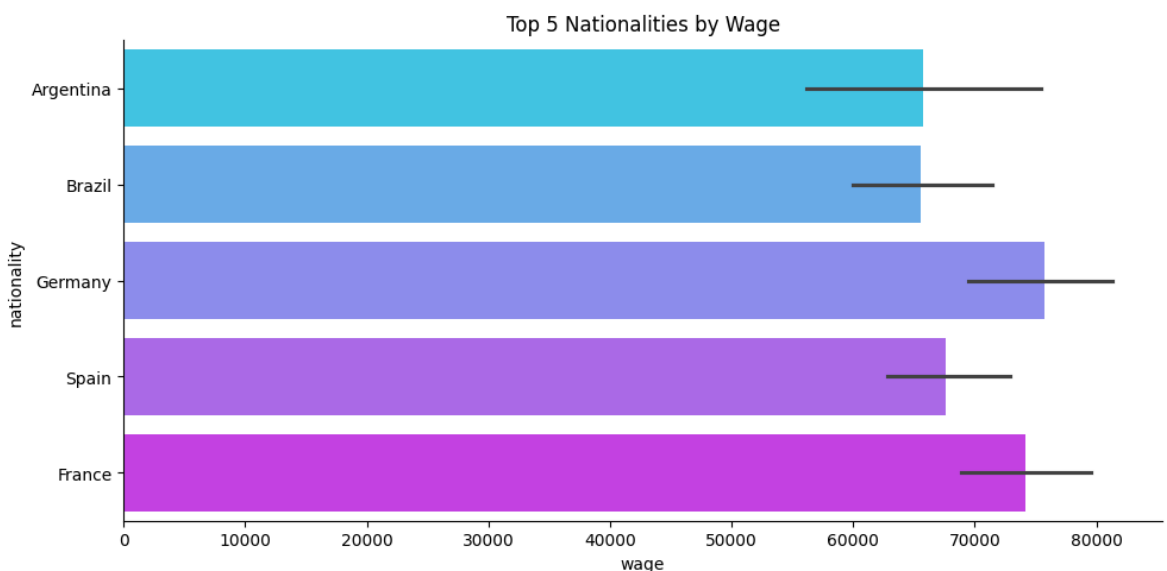


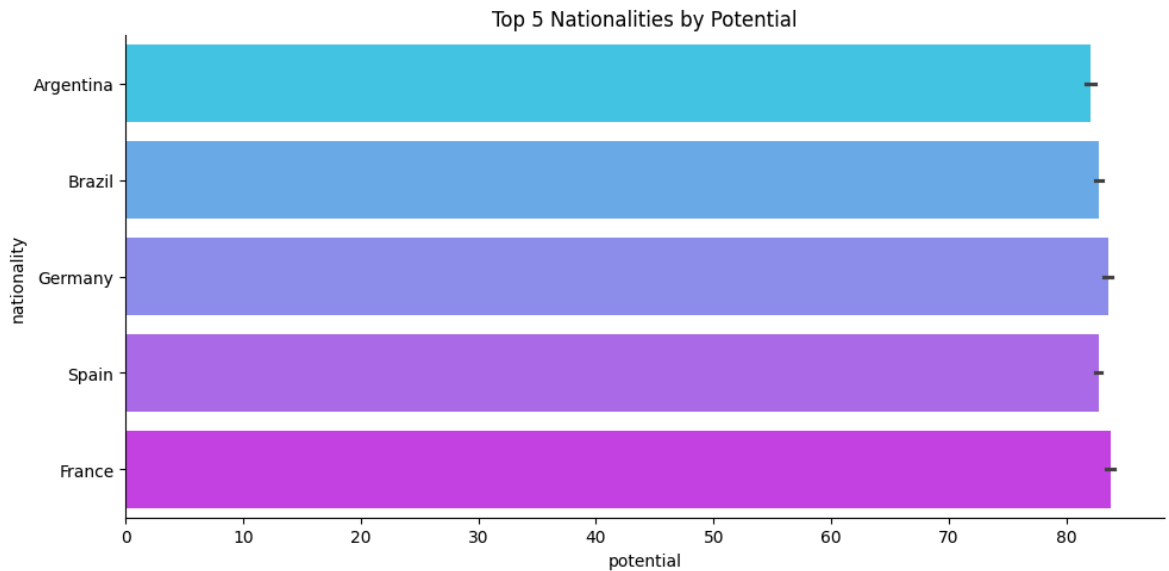
Nationality vs Wage and Potential

```
In [3]: top5_nations = df['nationality'].value_counts().head(5).index.tolist()
top5_df = df[df['nationality'].isin(top5_nations)]

sns.catplot(y='nationality', x='wage', data=top5_df, kind='bar', aspect=2
plt.title("Top 5 Nationalities by Wage")
plt.tight_layout()
plt.show()

sns.catplot(y='nationality', x='potential', data=top5_df, kind='bar', asp
plt.title("Top 5 Nationalities by Potential")
plt.tight_layout()
plt.show()
```



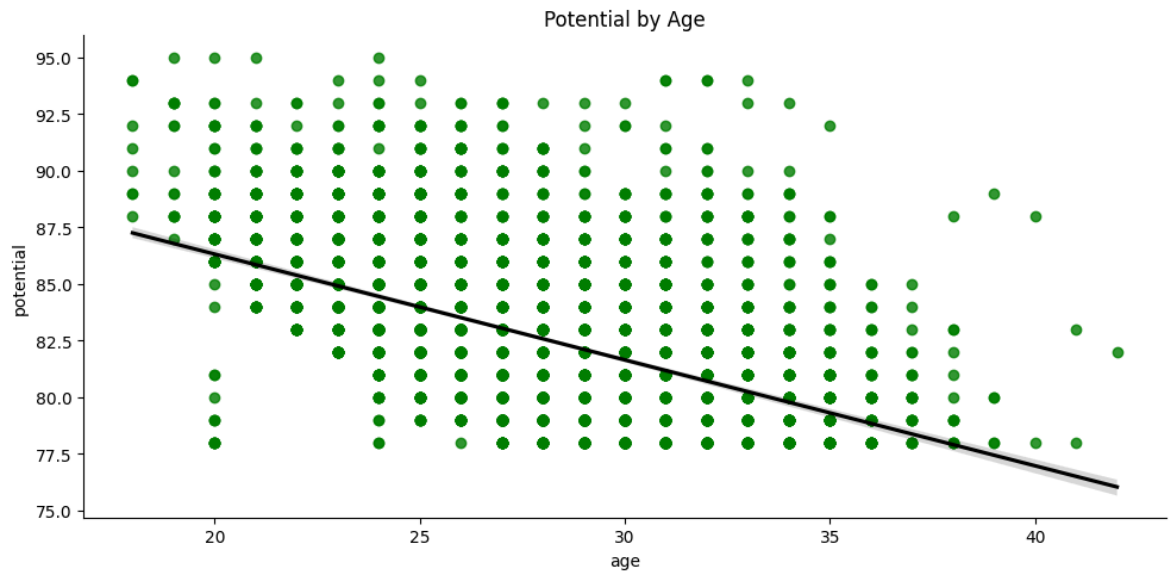


Age vs Wage and Potential

```
In [4]: sns.lmplot(x='age', y='wage', data=df, aspect=2,
                  scatter_kws={'color': 'blue'}, line_kws={'color': 'purple'})
plt.title("Wage by Age")
plt.tight_layout()
plt.show()

sns.lmplot(x='age', y='potential', data=df, aspect=2,
          scatter_kws={'color': 'green'}, line_kws={'color': 'black'})
plt.title("Potential by Age")
plt.tight_layout()
plt.show()
```

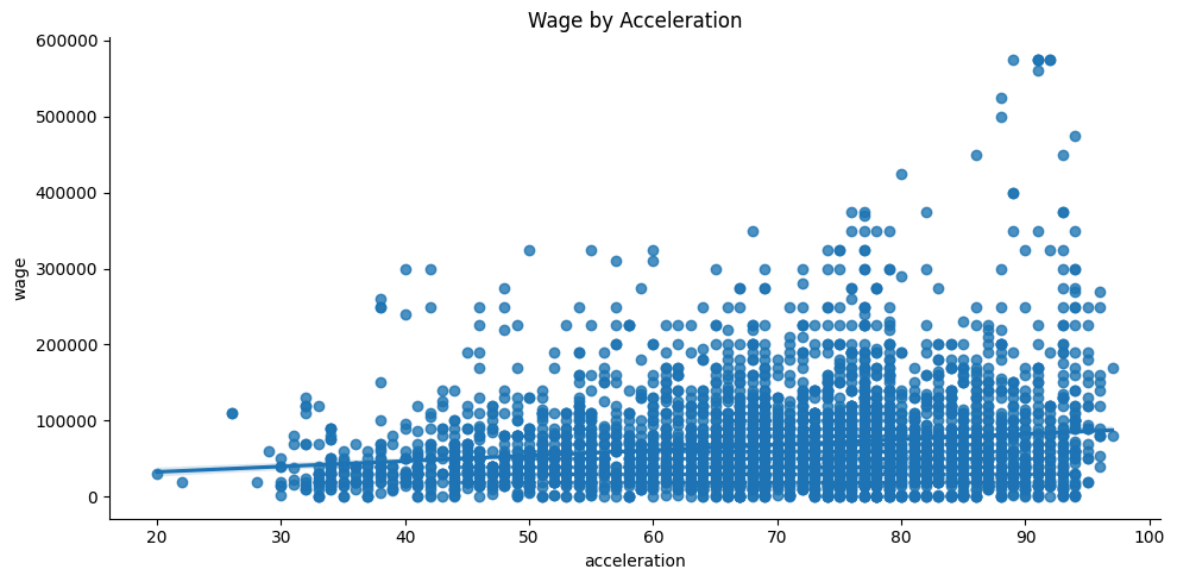


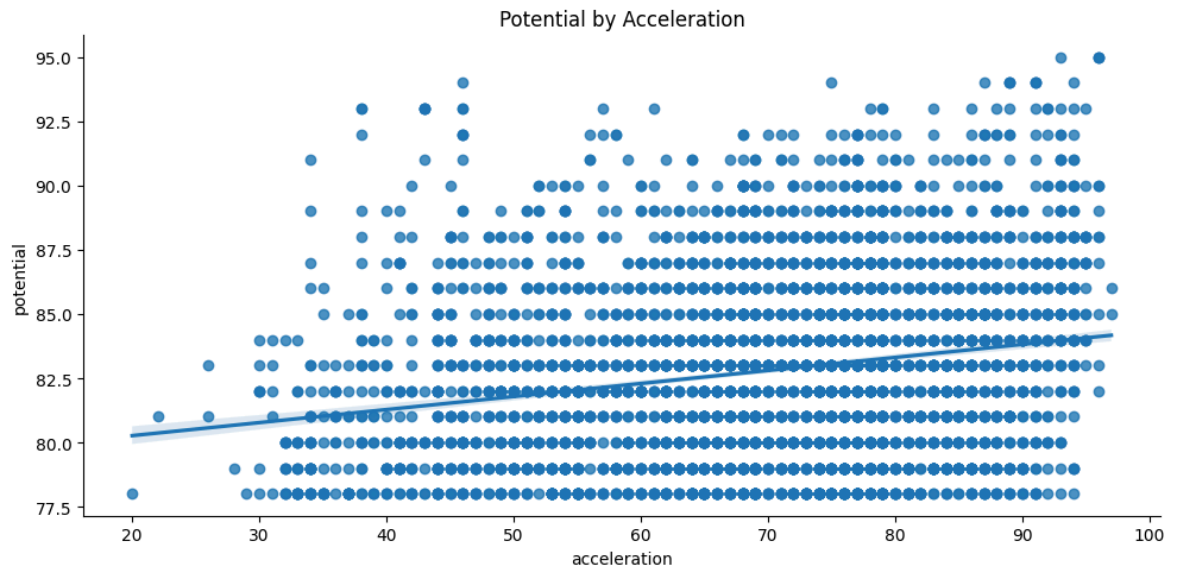


Acceleration vs Wage and Potential

```
In [5]: sns.lmplot(x='acceleration', y='wage', data=df, aspect=2)
plt.title("Wage by Acceleration")
plt.tight_layout()
plt.show()

sns.lmplot(x='acceleration', y='potential', data=df, aspect=2)
plt.title("Potential by Acceleration")
plt.tight_layout()
plt.show()
```

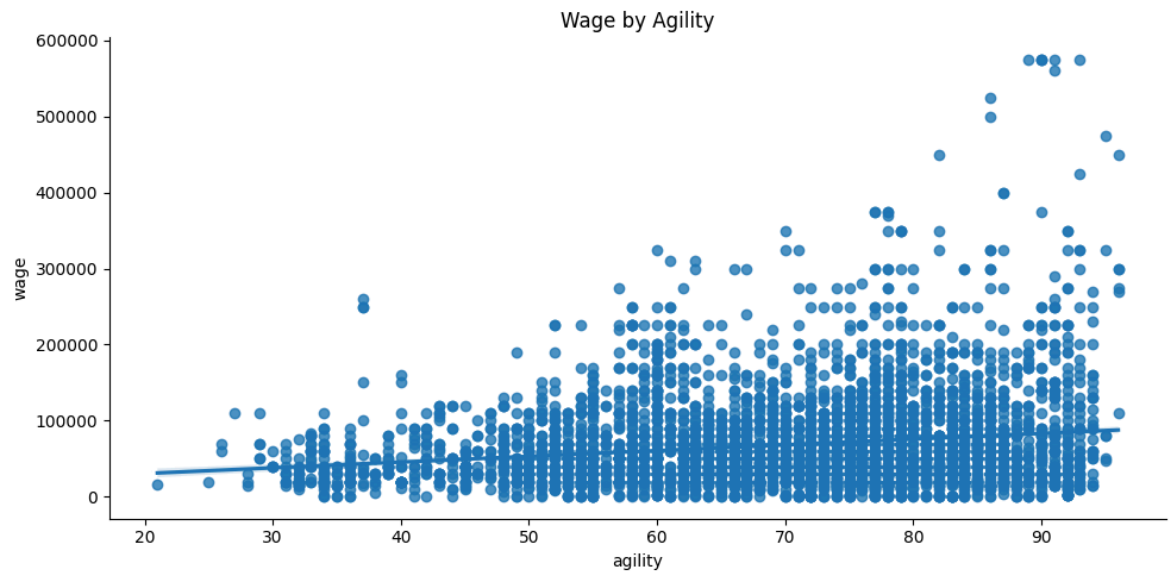


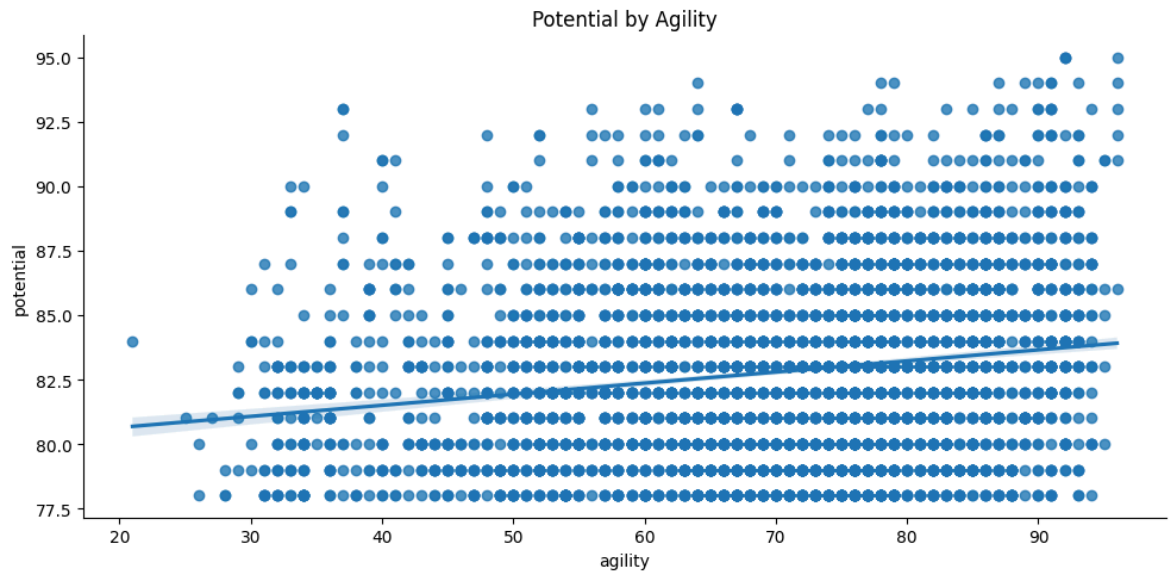


Agility vs Wage and Potential

```
In [6]: sns.lmplot(x='agility', y='wage', data=df, aspect=2)
plt.title("Wage by Agility")
plt.tight_layout()
plt.show()

sns.lmplot(x='agility', y='potential', data=df, aspect=2)
plt.title("Potential by Agility")
plt.tight_layout()
plt.show()
```



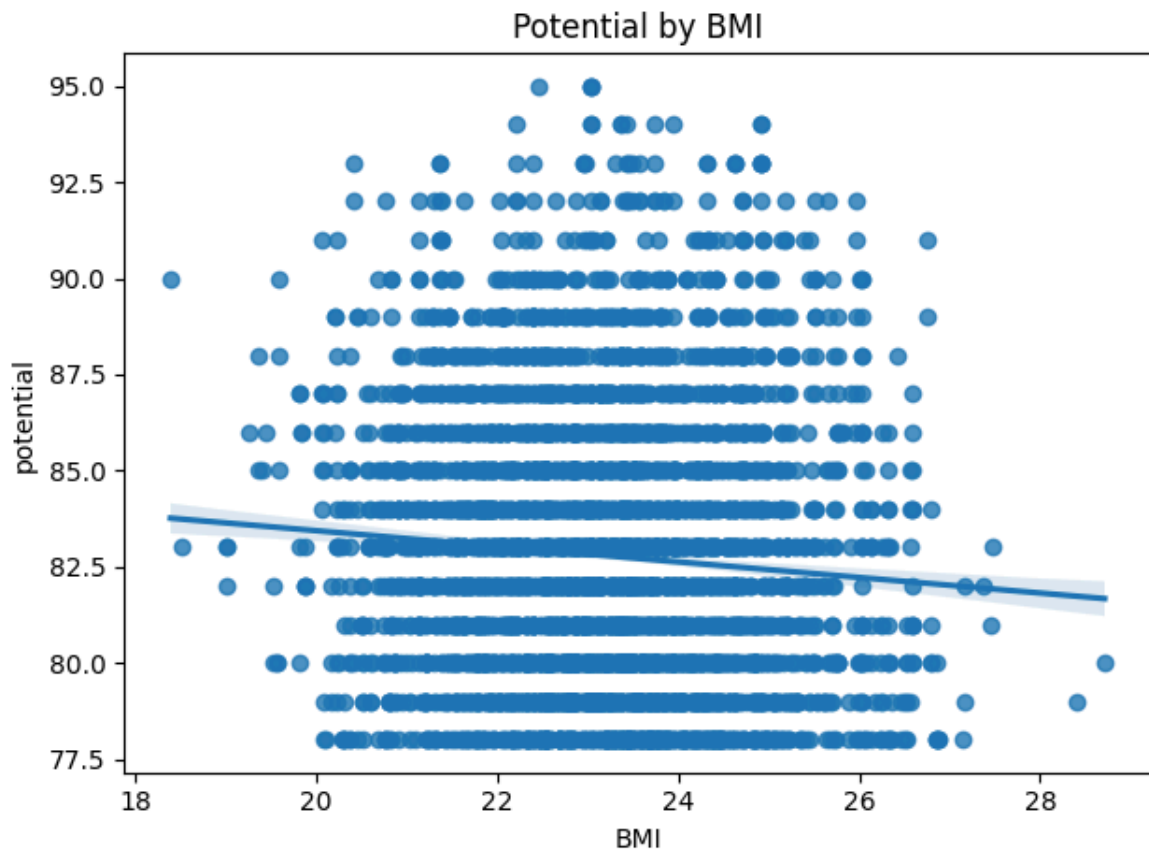


BMI vs Wage and Potential

```
In [7]: sns.regplot(x='BMI', y='wage', data=df)
plt.title("Wage by BMI")
plt.tight_layout()
plt.show()

sns.regplot(x='BMI', y='potential', data=df)
plt.title("Potential by BMI")
plt.tight_layout()
plt.show()
```





✓ Conclusion

- **Preferred foot** had minimal impact on wage or potential.
- **Nationality** influences both wage and potential. Argentina, Portugal, and Brazil dominate.
- **Age** shows players peak in potential between 26–30.
- **Acceleration** and **agility** significantly impact both wage and potential.
- **BMI** trends suggest leaner players earn more.

🎯 Hypothesis confirmed:

Top 5% FIFA 21 players **were faster** than those in FIFA 17.