```
library(tidyverse)
library(ggplot2)

nba_data <- read_csv("NBA_2021_2022_stats.csv")

#make data easier to read with fully typed out column names
NBA_22 <- nba_data %>%
  rename(
    Team = Tm,
    `Games Played` = G,
    `Offensive Rebounds Per Game` = ORB,
    `Defensive Rebounds Per Game` = DRB,
    `Assists Per Game` = AST,
    `3 Point Percentage` = `3P%`,
    `2 Point Percentage` = `2P%`,
    `Free Throw Percentage` = `FT%`,
    `Field Goal Percentage` = `FG%`,
    `Steals Per Game` = STL,
    `Blocks Per Game` = BLK,
    `Points Per Game` = PTS,
    Position = Pos
  )

#Average Fantasy Points (Arbitrary Scoring System used by
#NBA Fantasy Players in Points League Formats)
# -----------------Glossary----------------------------
# Points =                1.05 Fantasy Points
# Defensive Rebounds =    1.15 Fantasy Points
# Offensive Rebounds  =   1.25 Fantasy Points
# Steals =                2.00 Fantasy Points
# Blocks =                2.50 Fantasy Points
# Assists =               1.85 Fantasy Points

NBA_22 <- NBA_22 %>%
  mutate(
    `Average Fantasy Points` = (
      `Offensive Rebounds Per Game` * 1.25 +
        `Defensive Rebounds Per Game` * 1.15 +
        `Points Per Game` * 1.05 +
        `Blocks Per Game` * 2.5 +
        `Steals Per Game` * 2 +
        `Assists Per Game` * 1.85
    ),
    `Total Fantasy Points` = `Average Fantasy Points` * `Games Played`
  )

# Split Position into two different Columns if a player plays 2 Positions
# ##-## first ## is Position 1 and second ## is Position 2
# If not keep their position in Position 1 and Leave Position 2 blank

NBA_22 <- NBA_22 %>%
  separate(Position, into = c("Position 1", "Position 2"), sep = "-", fill = "right",
remove = FALSE)

# Drop the Position Column since I no longer need it
NBA_22 <- NBA_22 %>% select(-Position)


# Identifying starter level, young and prime players
# Defined by player 18 - 26
# (Rookies aged players to players entering their prime)
# Averaging 35 and Up fantasy points
rookie_to_prime_performers <- NBA_22 %>%
  filter(Age <= 26 & `Average Fantasy Points` >= 35)
```

```
count_performers <- rookie_to_prime_performers %>%
  group_by(`Position 1`) %>%
  summarize(Count = n())
# NBA Age vs Performance Analysis by Primary Position with Highlight on Young and Entering
Prime Players
# This script creates a scatter plot using the NBA_22 data set.
# The plot visualizes the relationship between a player's average fantasy points and their
age,
# with separate panels for each position. Additionally, it highlights young and prime
performers,
# defined as players aged 26 or younger with average fantasy points of 35 or more
ggplot(NBA_22, aes(y = Age, x = `Average Fantasy Points`)) +
  geom_point(aes(color = `Position 1`)) +
  geom_point(data = rookie_to_prime_performers, aes(y = Age, x = `Average Fantasy
Points`),
             color = 'black', size = 3, shape = 1) +
  facet_wrap(~ `Position 1`) +
  geom_text(data = count_performers, aes(label = Count, y = 18, x = 70), hjust = 1, vjust
= 1, size = 3, color = "blue") +
  scale_y_reverse(limits = c(max(NBA_22$Age), 18)) +
  scale_x_continuous(limits = c(0, 70)) +
  labs(title = 'Average Fantasy Points vs Age by Primary Position (Prime and Below)',
       y = 'Age',
       x = 'Average Fantasy Points') +
  theme_minimal()

# Identifying starter level veteran players
# Defined by Players Aged 30 and up
# Averaging 35 and up fantasy points
veteran_performers <- NBA_22 %>%
  filter(Age >= 30 & `Average Fantasy Points` >= 35)

count_performers <- veteran_performers %>%
  group_by(`Position 1`) %>%
  summarize(Count = n())
# NBA Age vs Performance Analysis by Primary Position with Highlight Veteran
# This script creates a scatter plot using the NBA_22 data set.
# The plot visualizes the relationship between a player's average fantasy points and their
age,
# with separate panels for each position. Additionally, it highlights young and prime
performers,
# defined as players aged 30 or older with average fantasy points of 35 or more

ggplot(NBA_22, aes(y = Age, x = `Average Fantasy Points`)) +
    geom_point(aes(color = `Position 1`)) +
  geom_point(data = veteran_performers, aes(y = Age, x = `Average Fantasy Points`),
             color = 'black', size = 3, shape = 1) +
  facet_wrap(~ `Position 1`) +
  geom_text(data = count_performers, aes(label = Count, y = 18, x = 70), hjust = 1, vjust
= 1, size = 3, color = "blue") +
  scale_y_reverse(limits = c(max(NBA_22$Age), 18)) +
  scale_x_continuous(limits = c(0, 70)) +
  labs(title = 'Average Fantasy Points vs Age by Primary Position (Longevity)',
       y = 'Age',
       x = 'Average Fantasy Points') +
  theme_minimal()

# Identifying Starter Level Players
# All Players Averaging 35 or more fantasy players
general_performers <- NBA_22 %>%
  filter(`Average Fantasy Points` >= 35)

count_performers <- general_performers %>%
  group_by(`Position 1`) %>%
```

```
  summarize(Count = n())
# NBA Age vs Performance Analysis by Primary Position with Highlights General
# The plot visualizes the relationship between a player's average fantasy points and their
age,
# with separate panels for each position. It highlights all starter level players
#, defined as players with average fantasy points of 35 or more.
ggplot(NBA_22, aes(y = Age, x = `Average Fantasy Points`)) +
  geom_point(aes(color = `Position 1`)) +
  geom_point(data = general_performers, aes(y = Age, x = `Average Fantasy Points`),
             color = 'black', size = 3, shape = 1) +
  facet_wrap(~ `Position 1`) +
  geom_text(data = count_performers, aes(label = Count, y = 18, x = 70), hjust = 1, vjust
= 1, size = 3, color = "blue") +
  scale_y_reverse(limits = c(max(NBA_22$Age), 18)) +
  scale_x_continuous(limits = c(0, 70)) +
  labs(title = 'Average Fantasy Points vs Age by Primary Position (General)',
       y = 'Age',
       x = 'Average Fantasy Points') +
  theme_minimal()


# Box Plot for 'Average Fantasy Points' by Position
# Used to Display Overall Fantasy Points Average Dominance by the
# PG position
ggplot(NBA_22, aes(x = `Position 1`, y = `Average Fantasy Points`, fill = `Position 1`)) +
  geom_boxplot() +
  labs(title = "Average Fantasy Points by Position", x = "Position", y = "Average Fantasy
Points") +
  theme_minimal() +
  theme(legend.position = "none")


# Create a density plot for 'Average Fantasy Points' for each position in 'Position 1'
# red mark at the 35 fantasy point to signify where starter level players start
ggplot(NBA_22, aes(x = `Average Fantasy Points`, fill = `Position 1`)) +
  geom_density(alpha = 0.7) +  # Adjust alpha for transparency
  geom_vline(xintercept = 35, color = "red", linetype = "dashed") +
  scale_fill_brewer(palette = "Set2") +
  labs(title = "Density Distribution of Average Fantasy Points by Position",
       x = "Average Fantasy Points",
       y = "Density",
       fill = "Position 1") +
  theme_minimal()

# a bar chart for the distribution of players across different positions
# shows how PGs are in the middle of the pack when it comes to total players
# yet stats show that theyre the best players to target in a draft
ggplot(NBA_22, aes(x = `Position 1`)) +
  geom_bar(fill = "blue", color = "black") +
  labs(title = "Distribution of Players Across Different Positions",
       x = "Position",
       y = "Number of Players") +
  theme_minimal()


# the faceted scatter plot
# shows how pgs are often young while playing alot of games but still
# producing fantasy points at an elite level
ggplot(NBA_22, aes(x = `Games Played`, y = `Total Fantasy Points`, color = Age)) +
  geom_point(alpha = 0.7) +
  scale_color_gradient(low = "red", high = "blue") +
  facet_wrap(~ `Position 1`, scales = 'free') +
  labs(title = "NBA Player Performance Analysis by Position",
       x = "Games Played",
       y = "Total Fantasy Points",
```

```
       color = "Age") +
  theme_minimal() +
  theme(legend.position = "bottom")

# Using aov function for ANOVA
anova_test <- aov(`Average Fantasy Points` ~ `Position 1`, data = NBA_22)

# Summary of the ANOVA test
anova_summary <- summary(anova_test)

capture.output(anova_summary, file = "anova results.txt")
```