SHAP_Analysis

2023-01-31

```r
library(tidyverse)
library(dplyr)
library(scales)
library(readxl)
All_Housing_Data <- read_csv("/Users/naimasagar/Green Housing Project/All_Housing_Dat
a.csv")
shap_data <- All_Housing_Data
```

Absolute Average Contribution - SHAP

```r
shap_data %>% select(-c("...1")) %>% select(-('Close Price':'Year Built')) %>%
  select_if(is.numeric) %>% abs() %>% summarise_all(mean) %>% t() %>%
  format(scientific = FALSE)
```

```
##                                    [,1]
## Green Building Verification_SHAP  "0.00001647564"
## Green Energy Count_SHAP           "0.00006822836"
## Green Energy Efficient_SHAP       "0.00269516335"
## Green Sustainability_SHAP         "0.00010242689"
## Accessibility Features_SHAP       "0.00047976036"
## Bedroom Count_SHAP                "0.00396511610"
## Garage Space Count_SHAP           "0.00570066524"
## Homeowner Association_SHAP        "0.01335256374"
## Latitude_SHAP                     "0.07172512260"
## Longitude_SHAP                    "0.05700293799"
## Occupant Type_SHAP                "0.00379965125"
## Off Market Month_SHAP             "0.00973766624"
## Off Market Year_SHAP              "0.07099823013"
## Patio & Porch Features_SHAP       "0.00177347998"
## Private Pool_SHAP                 "0.01305626529"
## Property Condition at Sale_SHAP   "0.01472112614"
## Security Features_SHAP            "0.00190645802"
## Square Feet_SHAP                  "0.09446517517"
## View_SHAP                         "0.00742329337"
## Year Built_SHAP                   "0.01950729931"
```
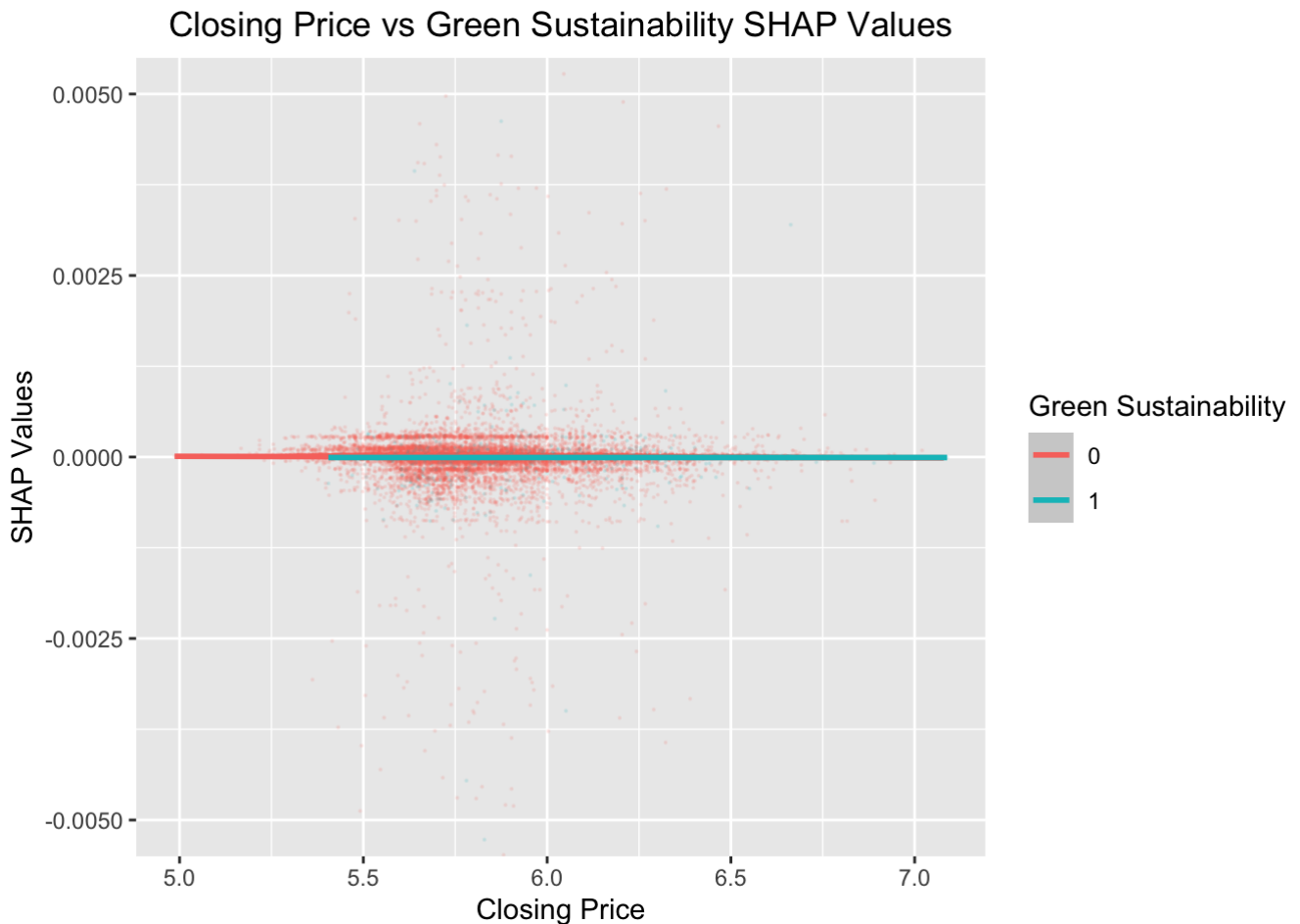
Closing Price & Green Energy Efficient SHAP Values

```
shap_data %>% select ('Close Price', 'Green Energy Efficient_SHAP', 'Green Energy Eff
icient') %>% na.omit() %>%
  rename(price='Close Price', efficient_SHAP='Green Energy Efficient_SHAP', green_ene
rgy='Green Energy Efficient') %>%
  mutate(green_energy = as.character(green_energy)) %>%
  ggplot(aes(x = price, y = efficient_SHAP, color = green_energy)) + geom_point(alpha
= 0.1, size = 0.1) + geom_smooth(size = 1) +
  coord_cartesian(ylim = c(-0.005, 0.005)) +
  labs(title="Closing Price vs Green Energy Efficient SHAP Values", x="Closing Pric
e", y="SHAP Values", color="Green Energy Efficient") +
  theme(plot.title = element_text(hjust = 0.5))
```
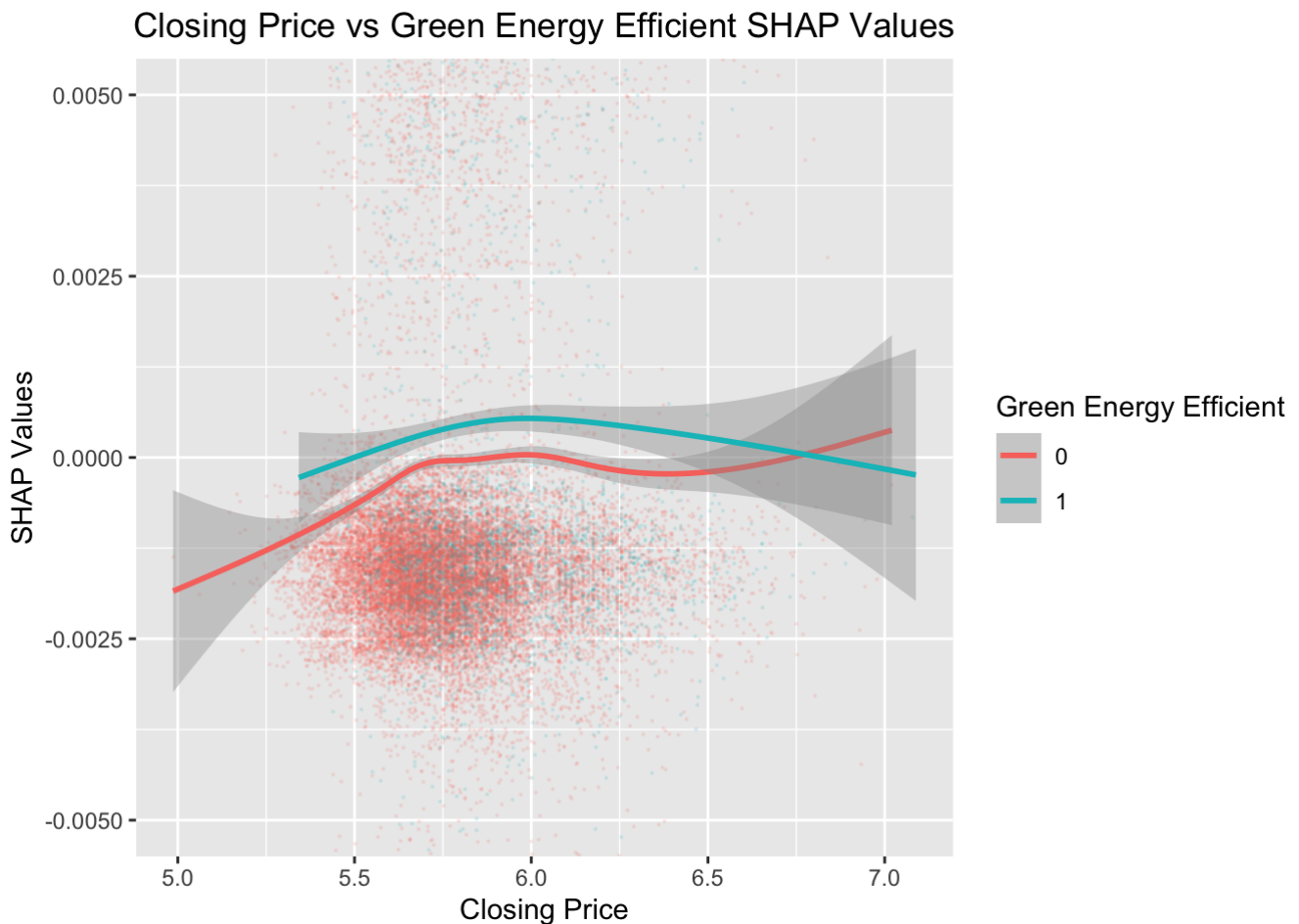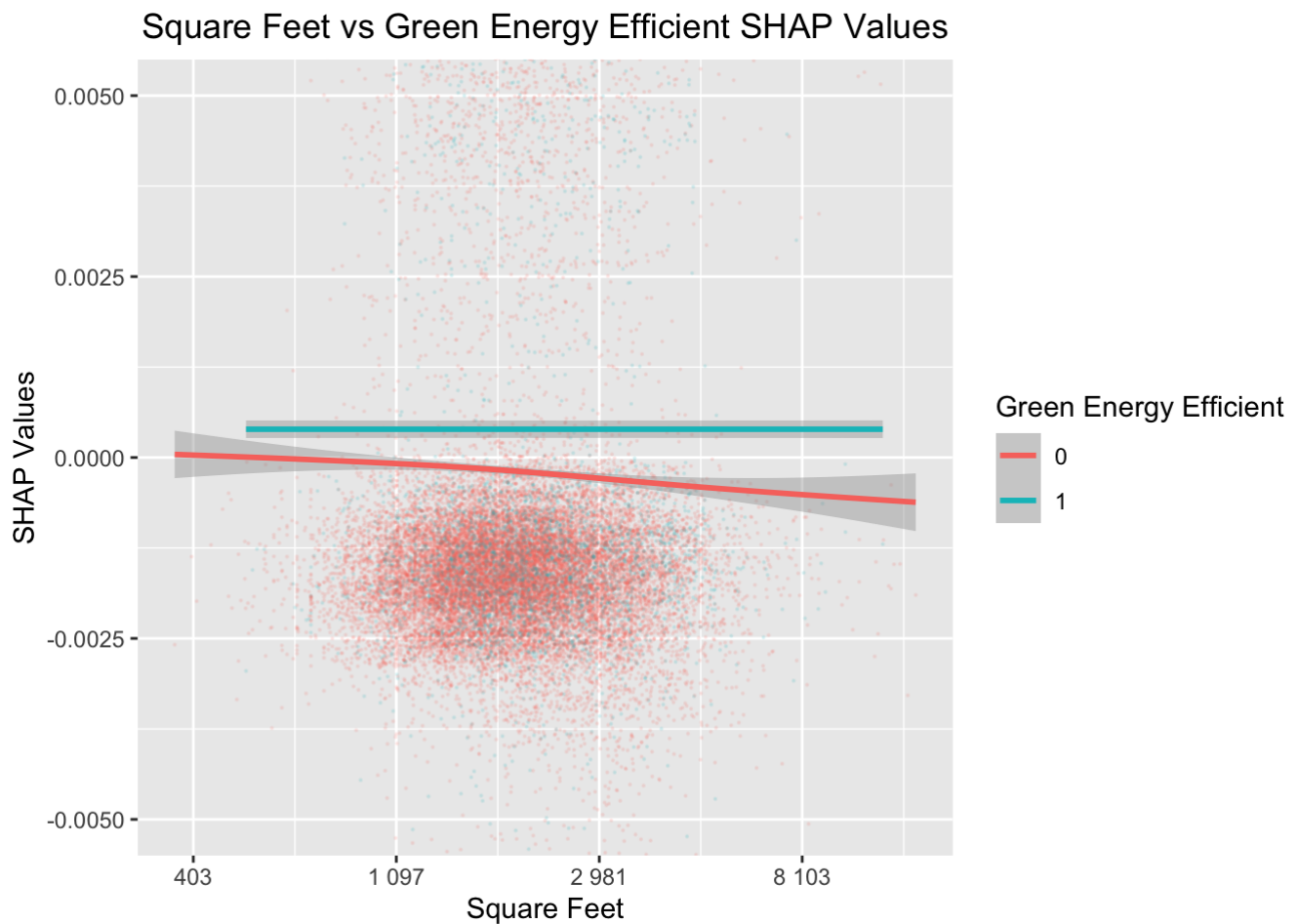

Closing Price & Green Sustainability SHAP Values

```
shap_data %>% select ('Close Price', 'Green Sustainability_SHAP', 'Green Sustainabili
ty') %>% na.omit() %>%
  rename(price='Close Price', sustainability_SHAP='Green Sustainability_SHAP', green_
sustainability='Green Sustainability') %>%
  mutate(green_sustainability = as.character(green_sustainability)) %>%
  ggplot(aes(x = price, y = sustainability_SHAP, color = green_sustainability)) + geo
m_point(alpha = 0.1, size = 0.1) +
  geom_smooth(size = 1) + coord_cartesian(ylim = c(-0.005, 0.005)) +
  labs(title="Closing Price vs Green Sustainability SHAP Values", x="Closing Price",
y="SHAP Values", color="Green Sustainability") +
  theme(plot.title = element_text(hjust = 0.5))
```
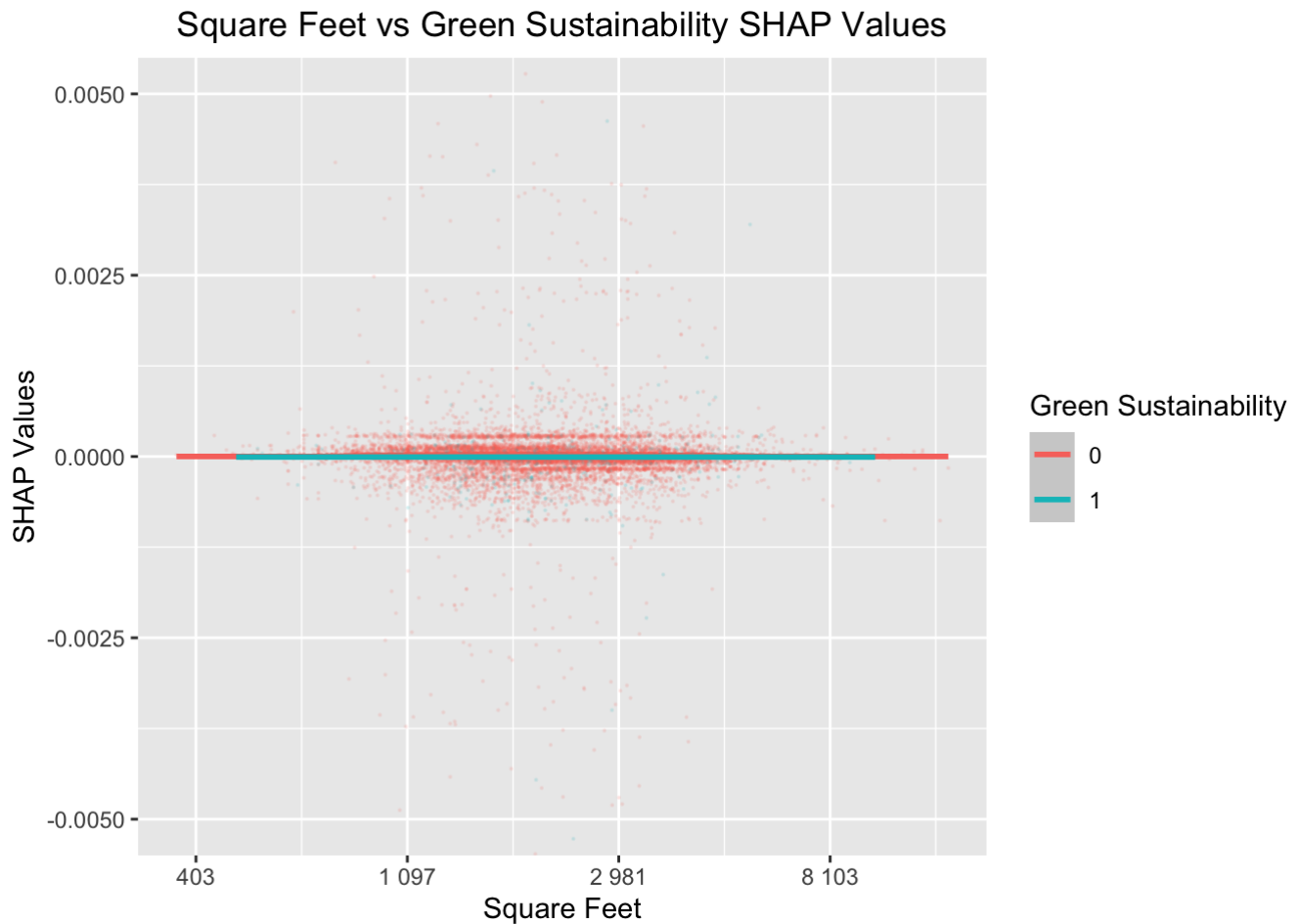

Square Feet & Green Energy Efficient SHAP Values

```
shap_data %>% select ('Square Feet', 'Green Energy Efficient_SHAP', 'Green Energy Eff
icient') %>% na.omit() %>%
   rename(sqft='Square Feet', efficient_SHAP='Green Energy Efficient_SHAP', green_ener
gy='Green Energy Efficient') %>%
   mutate(green_energy = as.character(green_energy)) %>%
   ggplot(aes(x = sqft, y = efficient_SHAP, color = green_energy)) + geom_point(alpha
= 0.1, size = 0.1) + geom_smooth(size = 1) +
   scale_x_continuous(trans='log', labels = label_number(accuracy = 1)) + coord_cartes
ian(ylim = c(-0.005, 0.005)) +
   labs(title="Square Feet vs Green Energy Efficient SHAP Values", x="Square Feet", y
="SHAP Values", color="Green Energy Efficient") +
   theme(plot.title = element_text(hjust = 0.5))
```


Square Feet & Green Sustainability SHAP Values

```
shap_data %>% select ('Square Feet', 'Green Sustainability_SHAP', 'Green Sustainabili
ty') %>% na.omit() %>%
  rename(sqft='Square Feet', sustainability_SHAP='Green Sustainability_SHAP', green_s
ustainability='Green Sustainability') %>%
  mutate(green_sustainability = as.character(green_sustainability)) %>%
  ggplot(aes(x = sqft, y = sustainability_SHAP, color = green_sustainability)) + geom
_point(alpha = 0.1, size = 0.1) +
  geom_smooth(size = 1) + scale_x_continuous(trans='log', labels = label_number(accur
acy = 1)) + coord_cartesian(ylim = c(-0.005, 0.005)) +
  labs(title="Square Feet vs Green Sustainability SHAP Values", x="Square Feet", y="S
HAP Values", color="Green Sustainability") +
  theme(plot.title = element_text(hjust = 0.5))
```


Year Built & Green Energy Efficient SHAP Values

```
shap_data %>% select ('Year Built', 'Green Energy Efficient_SHAP', 'Green Energy Effi
cient') %>% na.omit() %>%
  rename(year_built='Year Built', efficient_SHAP='Green Energy Efficient_SHAP', green
_energy='Green Energy Efficient') %>%
  mutate(green_energy = as.character(green_energy)) %>%
  ggplot(aes(x = year_built, y = efficient_SHAP, color = green_energy)) + geom_point
(alpha = 0.1, size = 0.1) + geom_smooth(size = 1) +
  scale_x_continuous(n.breaks = 10, labels = label_number(accuracy = 1)) + coord_cart
esian(ylim = c(-0.005, 0.005)) +
  labs(title="Year Built vs Green Energy Efficient SHAP Values", x="Year Built", y="S
HAP Values", color="Green Energy Efficient") +
  theme(plot.title = element_text(hjust = 0.5))
```
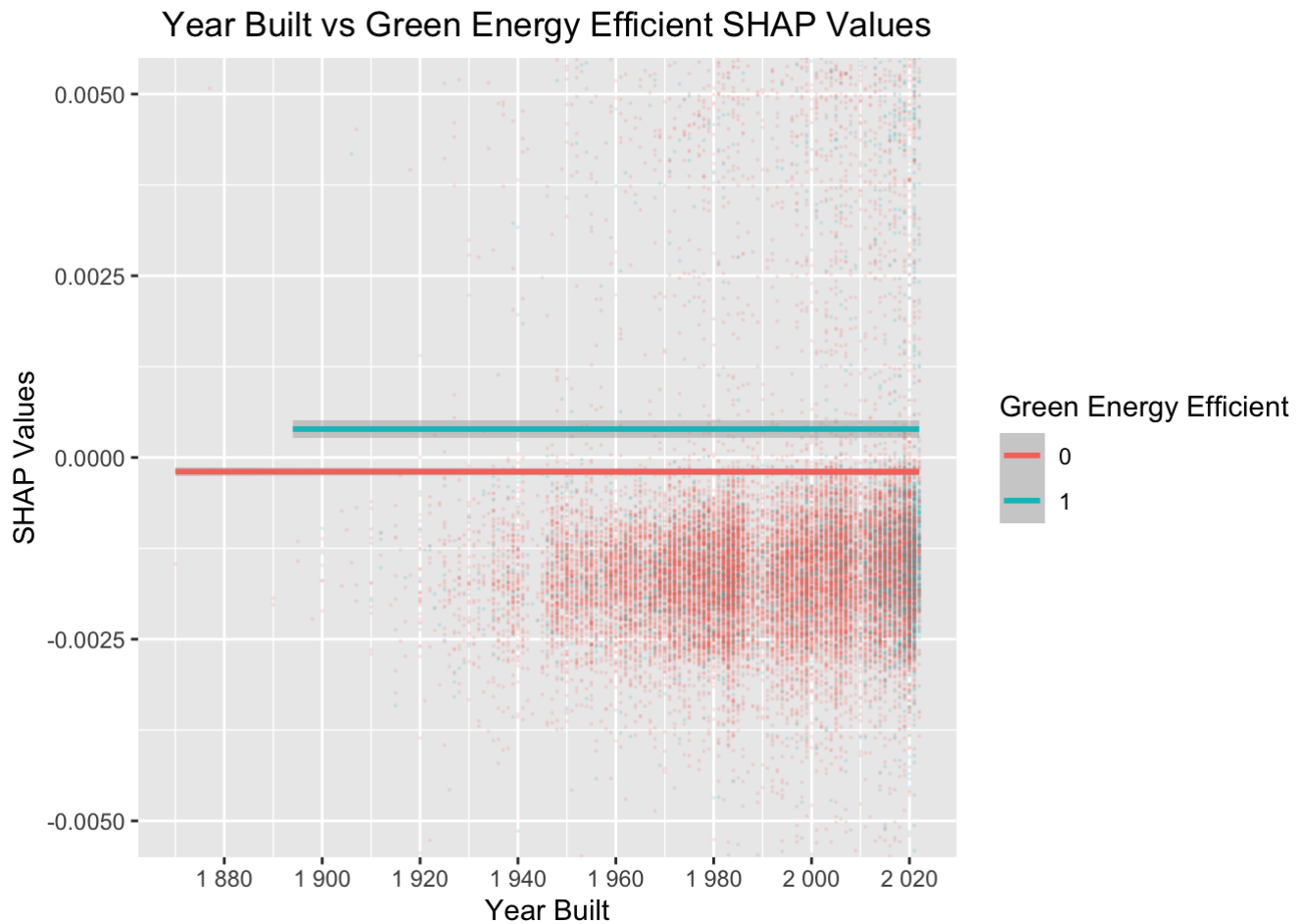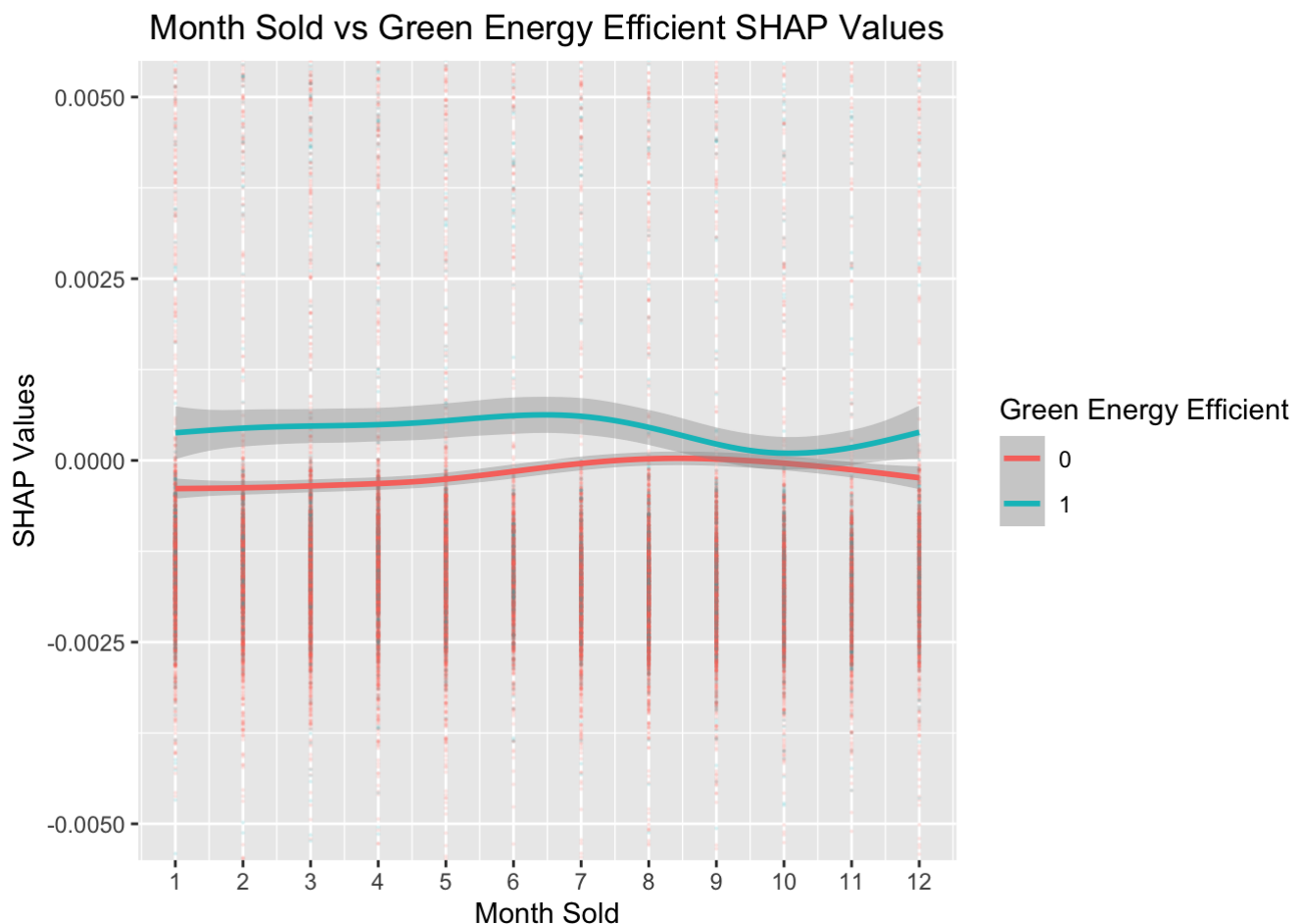

Year Built & Green Sustainability SHAP Values

```
shap_data %>% select ('Year Built', 'Green Sustainability_SHAP', 'Green Sustainabilit
y') %>% na.omit() %>%
  rename(year_built='Year Built', sustainability_SHAP='Green Sustainability_SHAP', gr
een_sustainability='Green Sustainability') %>%
  mutate(green_sustainability = as.character(green_sustainability)) %>%
  ggplot(aes(x = year_built, y = sustainability_SHAP, color = green_sustainability))
+ geom_point(alpha = 0.1, size = 0.1) +
  geom_smooth(size = 1) + scale_x_continuous(n.breaks=10, labels=label_number(accurac
y = 1)) + coord_cartesian(ylim = c(-0.005, 0.005)) +
  labs(title="Year Built vs Green Sustainability SHAP Values", x="Year Built", y="SHA
P Values", color="Green Sustainability") +
  theme(plot.title = element_text(hjust = 0.5))
```
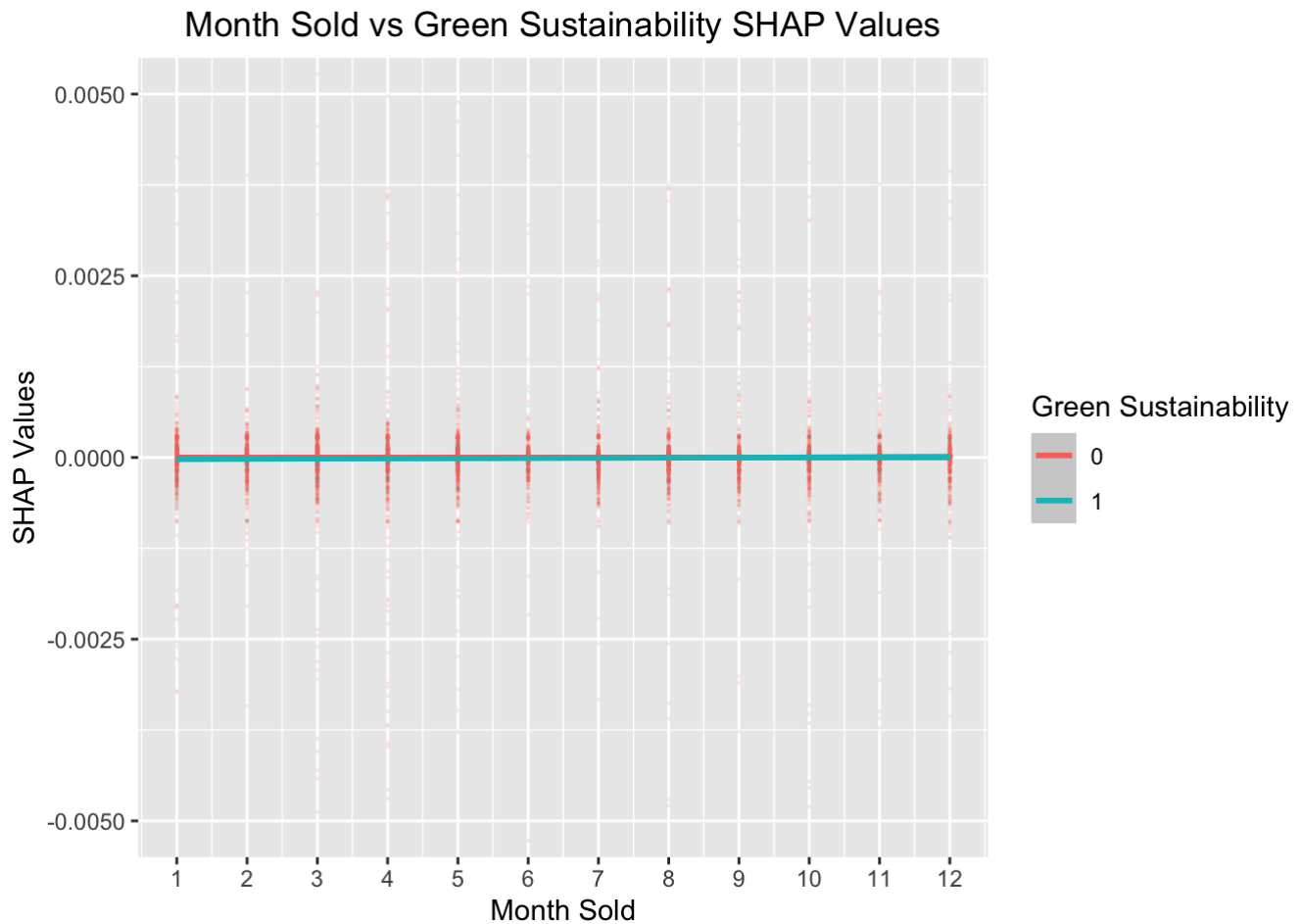

Year Built vs Green Sustainability SHAP Values

Month Sold & Green Energy Efficient SHAP Values

```
shap_data %>% select ('Off Market Month', 'Green Energy Efficient_SHAP', 'Green Energ
y Efficient') %>% na.omit() %>%
  rename(month_sold='Off Market Month', efficient_SHAP='Green Energy Efficient_SHAP',
green_energy='Green Energy Efficient') %>%
  mutate(green_energy = as.character(green_energy)) %>%
  ggplot(aes(x = month_sold, y = efficient_SHAP, color = green_energy)) + geom_point
(alpha = 0.1, size = 0.1) + geom_smooth(size = 1) +
  scale_x_continuous(n.breaks = 12, labels = label_number(accuracy = 1)) + coord_cart
esian(ylim = c(-0.005, 0.005)) +
  labs(title="Month Sold vs Green Energy Efficient SHAP Values", x="Month Sold", y="S
HAP Values", color="Green Energy Efficient") +
  theme(plot.title = element_text(hjust = 0.5))
```


Month Sold & Green Sustainability SHAP Values

```
shap_data %>% select ('Off Market Month', 'Green Sustainability_SHAP', 'Green Sustain
ability') %>% na.omit() %>%
  rename(month_sold='Off Market Month', sustainability_SHAP='Green Sustainability_SHA
P', green_sustainability='Green Sustainability') %>%
  mutate(green_sustainability = as.character(green_sustainability)) %>%
  ggplot(aes(x = month_sold, y = sustainability_SHAP, color = green_sustainability))
+ geom_point(alpha = 0.1, size = 0.1) +
  geom_smooth(size = 1) + scale_x_continuous(n.breaks=12, labels=label_number(accurac
y = 1)) + coord_cartesian(ylim = c(-0.005, 0.005)) +
  labs(title = "Month Sold vs Green Sustainability SHAP Values", x = "Month Sold", y
= "SHAP Values", color = "Green Sustainability") +
  theme(plot.title = element_text(hjust = 0.5))
```


Year Sold & Green Energy Efficient SHAP Values

```
shap_data %>% select ('Off Market Year', 'Green Energy Efficient_SHAP', 'Green Energy
Efficient') %>% na.omit() %>%
  rename(year_sold='Off Market Year', efficient_SHAP='Green Energy Efficient_SHAP', g
reen_energy='Green Energy Efficient') %>%
  mutate(year_sold = as.character(year_sold)) %>% mutate(green_energy = as.character
(green_energy)) %>% group_by(green_energy) %>%
  ggplot(aes(x = year_sold, y = efficient_SHAP, color = green_energy)) + geom_boxplot
() +
  labs(title = "Year Sold vs Green Energy Efficient SHAP Values", x = "Year Sold", y
= "SHAP Values", color = "Green Energy Efficient") +
  theme(plot.title = element_text(hjust = 0.5))
```


Year Sold vs Green Energy Efficient SHAP Values