



# Forest Cover Type Prediction

Pratik Barjatiya

[pratikbarjatiya@gmail.com](mailto:pratikbarjatiya@gmail.com)



## DataSet:

### Dataset info

Number of variables	56
Number of observations	15120
Missing cells	0 (0.0%)
Duplicate rows	0 (0.0%)
Total size in memory	6.5 MiB
Average record size in memory	448.0 B

### Variables types

Numeric	12
Categorical	0
Boolean	42
Date	0
URL	0
Text (Unique)	0
Rejected	2
Unsupported	0

### Variables:

- Elevation - Elevation in meters
- Aspect - Aspect in degrees azimuth
- Slope - Slope in degrees
- Horizontal\_Distance\_To\_Hydrology - Horizontal Distance to nearest surface water features
- Vertical\_Distance\_To\_Hydrology - Vert Dist to nearest surface water features
- Horizontal\_Distance\_To\_Roadways - Horizontal Distance to nearest roadway
- Hillshade\_9am (0 to 255 index) - Hill shade index at 9am, summer solstice
- Hillshade\_Noon (0 to 255 index) - Hill shade index at noon, summer solstice
- Hillshade\_3pm (0 to 255 index) - Hill shade index at 3pm, summer solstice
- Horizontal\_Distance\_To\_Fire\_Points - Horizontal Distance to nearest wildfire ignition points
- Wilderness\_Area (4 binary columns, 0 = absence or 1 = presence) - Wilderness area designation
- Soil\_Type (40 binary columns, 0 = absence or 1 = presence) - Soil Type designation
- Cover\_Type (7 types, integers 1 to 7) - Forest Cover Type designation

# Correlation:

Hillshade\_9am and Hillshade\_3pm = -0.78

Horizontal\_Distance & Vertical\_Distance To\_Hydrology = 0.65

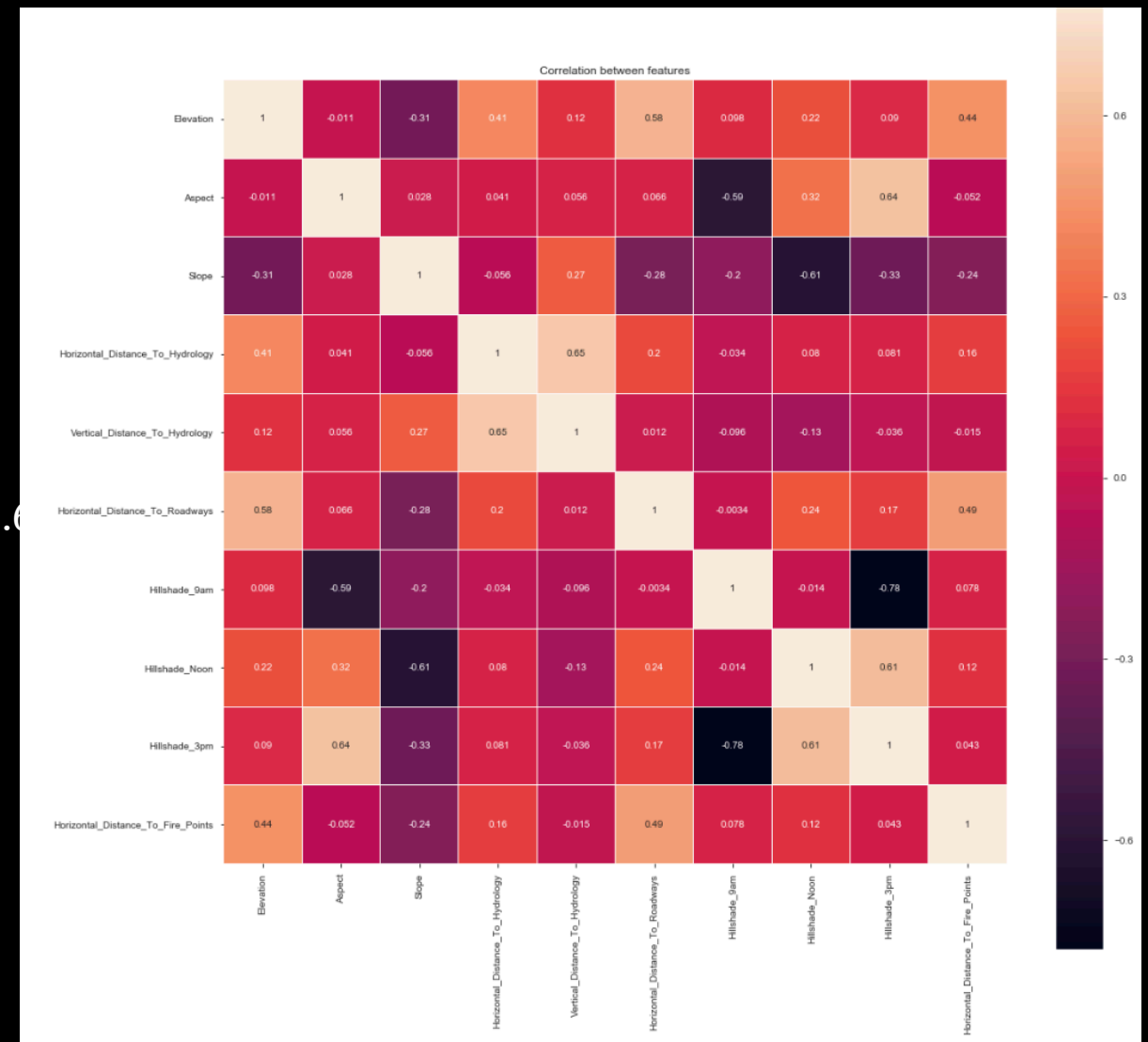
Aspect and Hillshade\_3pm = 0.64

Hillshade\_Noon and Hillshade\_3pm = 0.61

Slope and Hillshade\_Noon = -0.61

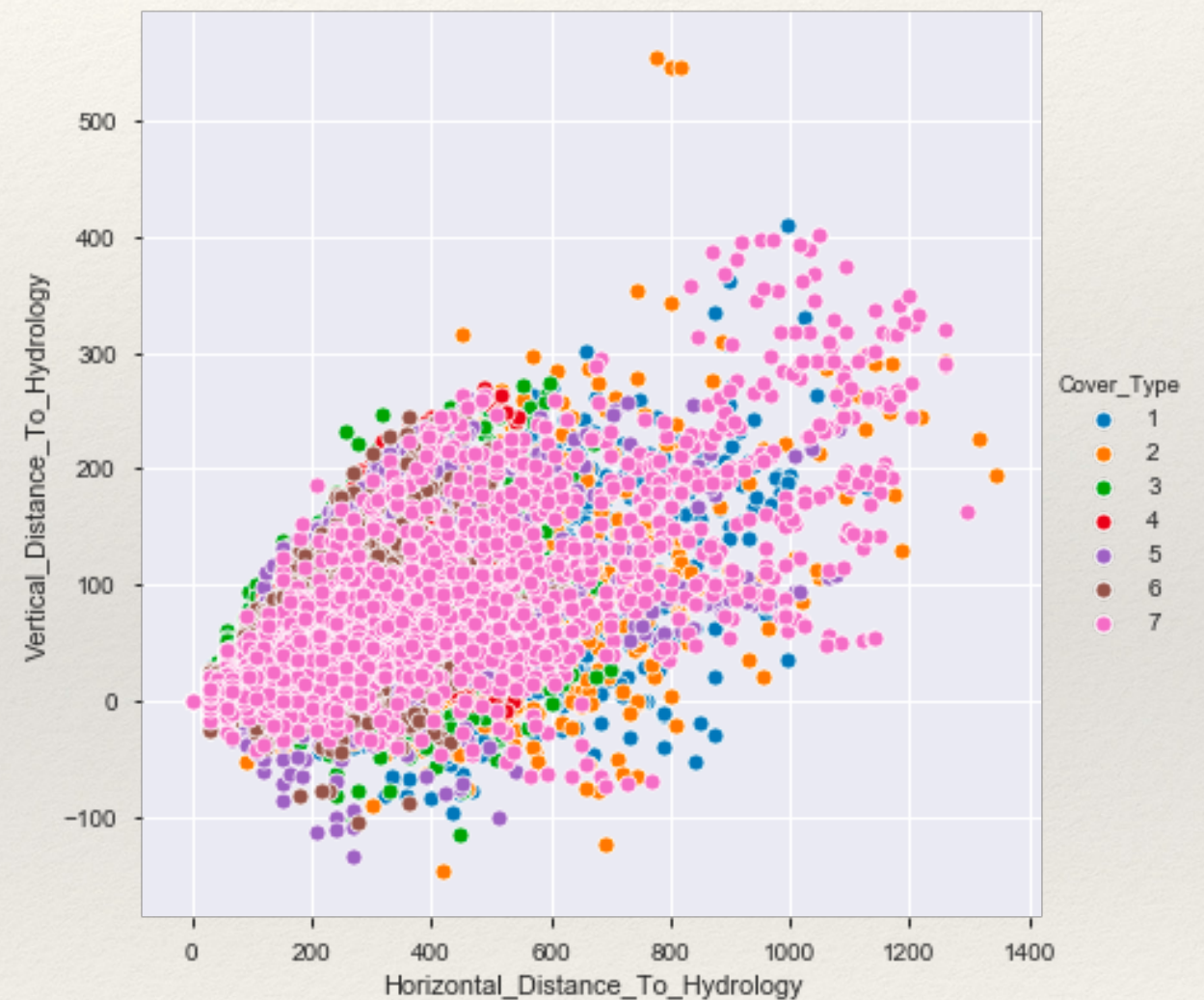
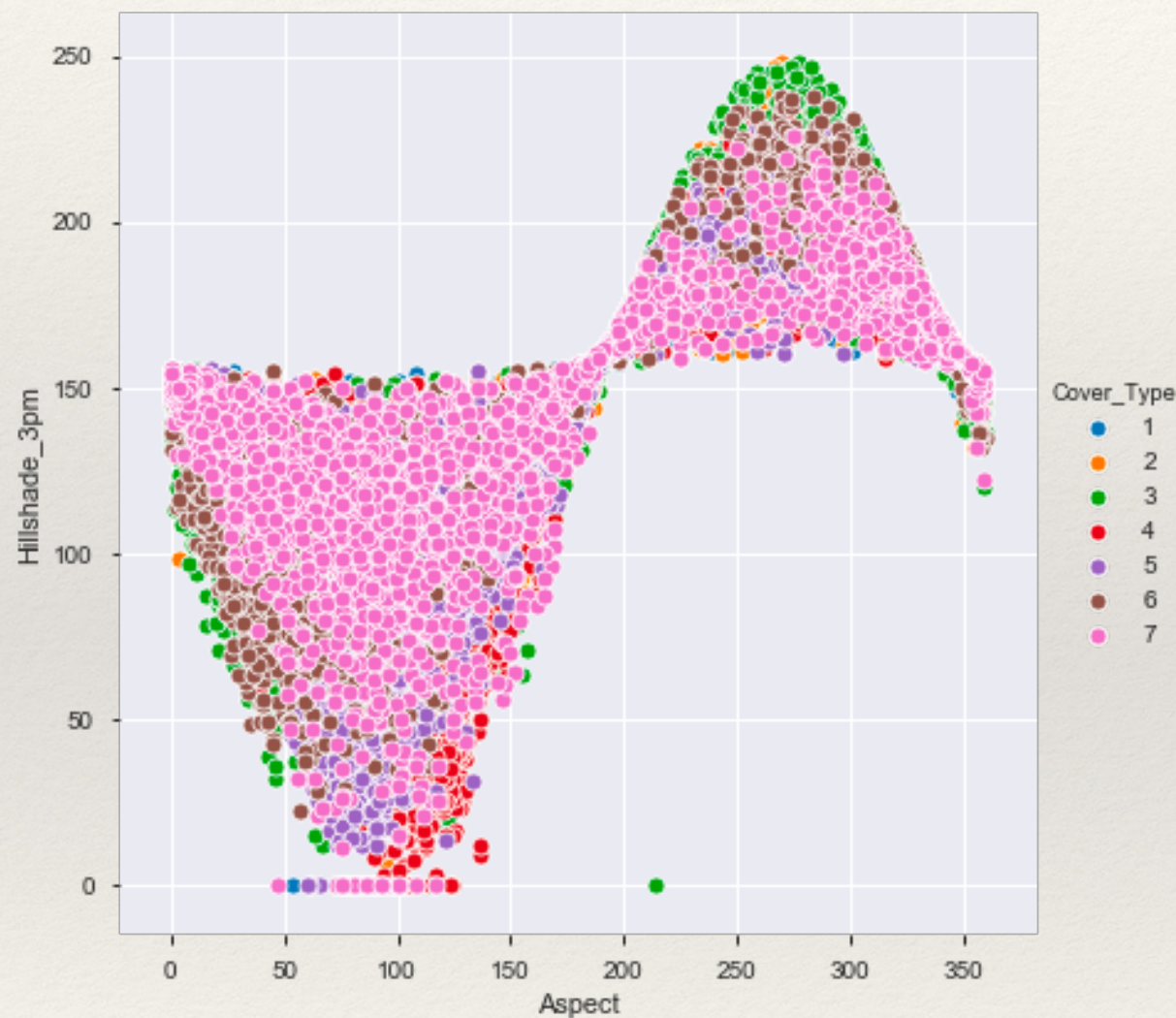
Aspect and Hillshade\_9am = -0.59

Elevation and Horizontal\_Distance\_To\_Roadways = 0.58



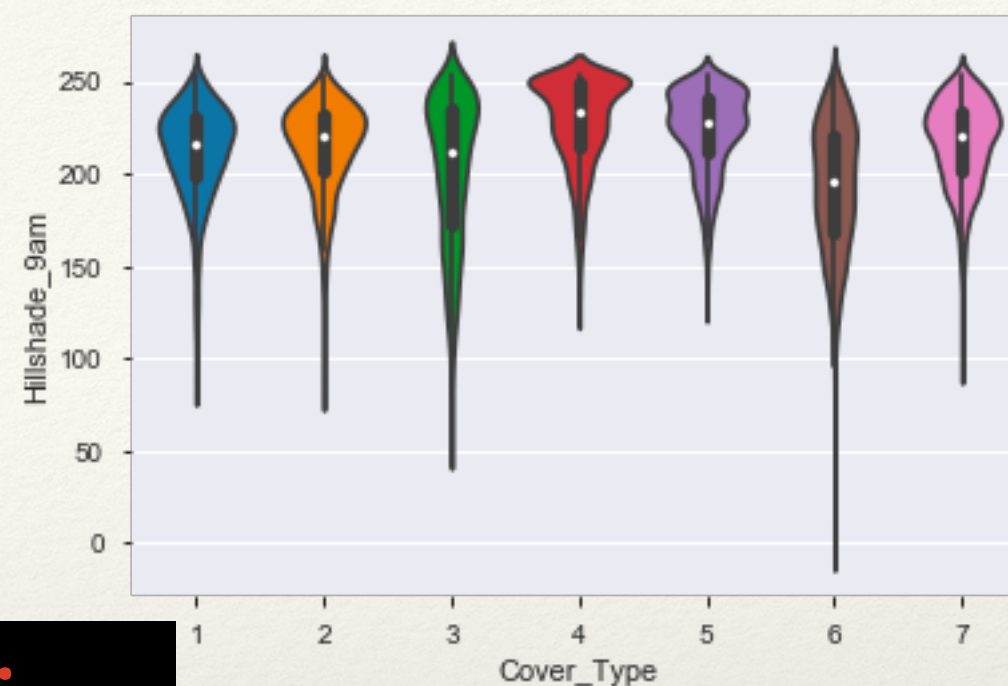
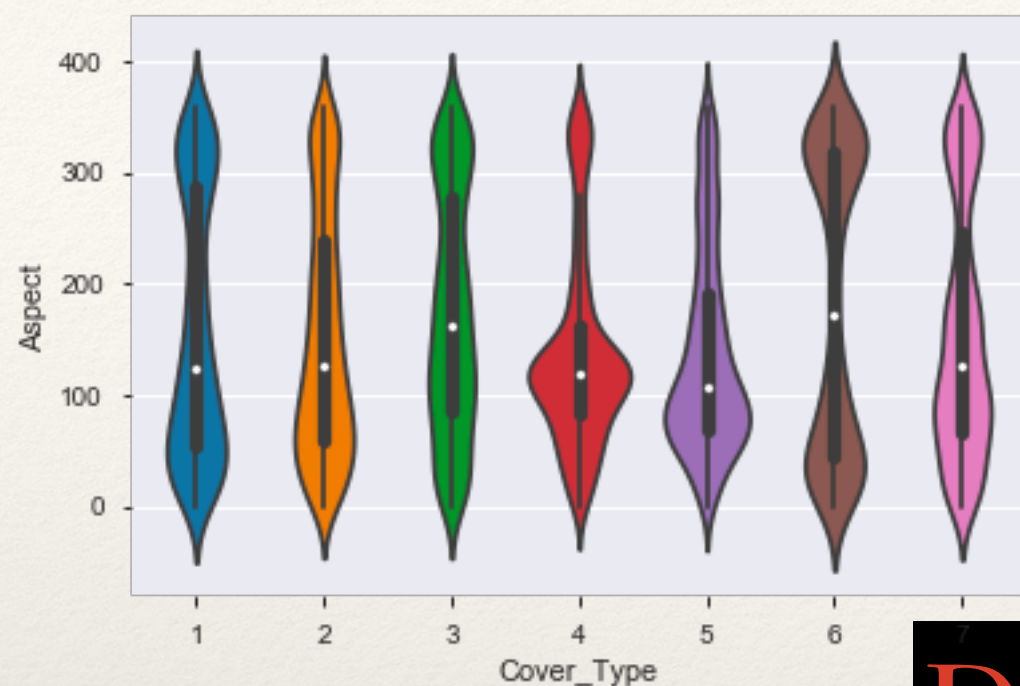


Horizontal and vertical distance to hydrology seems to have a linear relation  
Hillside and Aspect seems to have a sigmoid relation

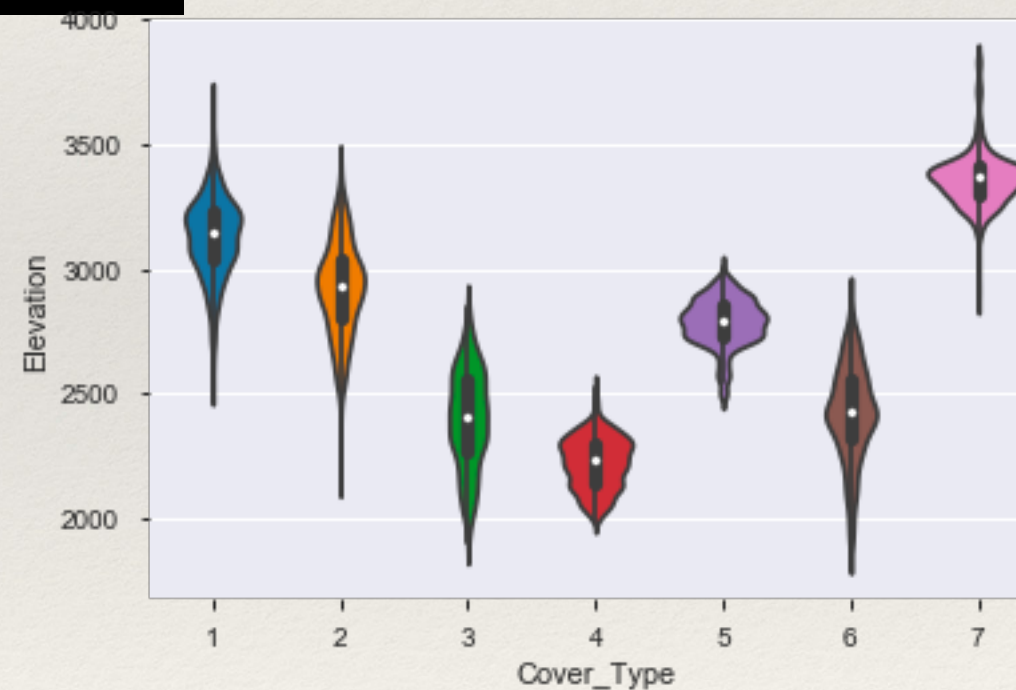
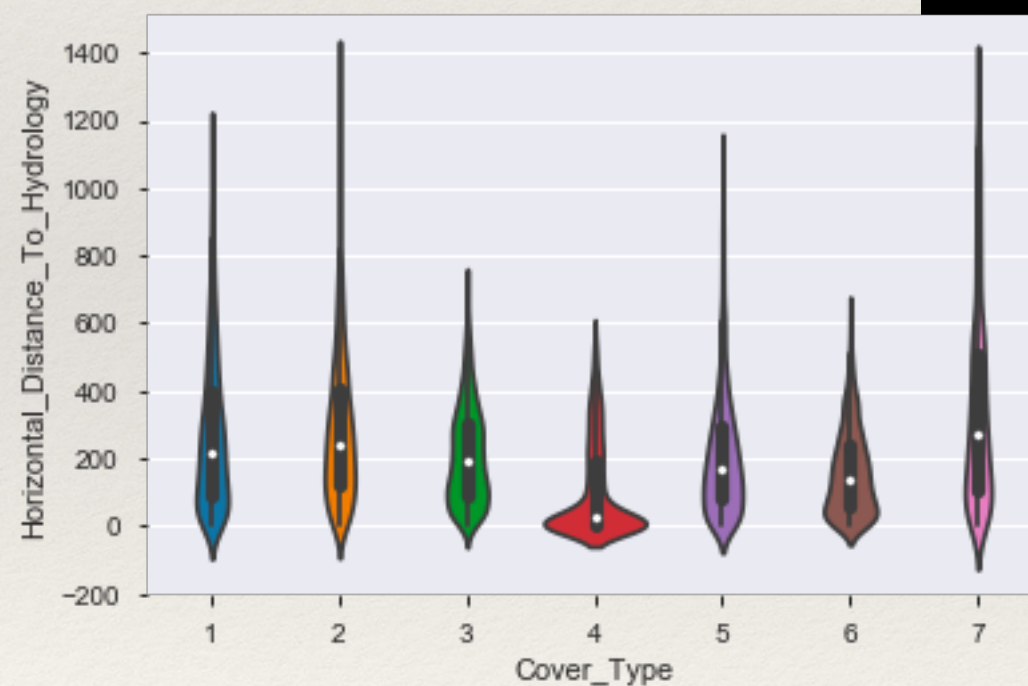


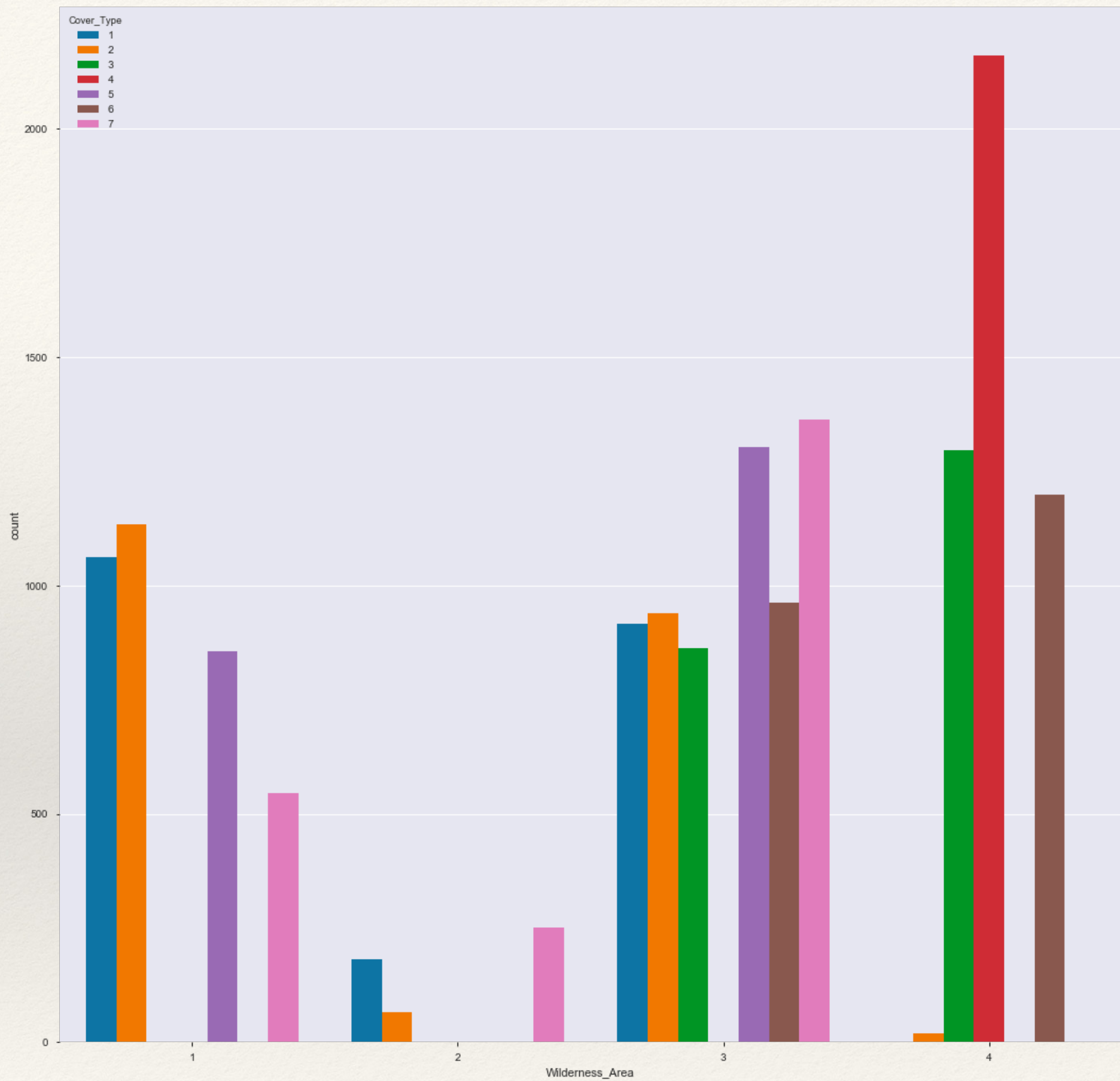
Distribution





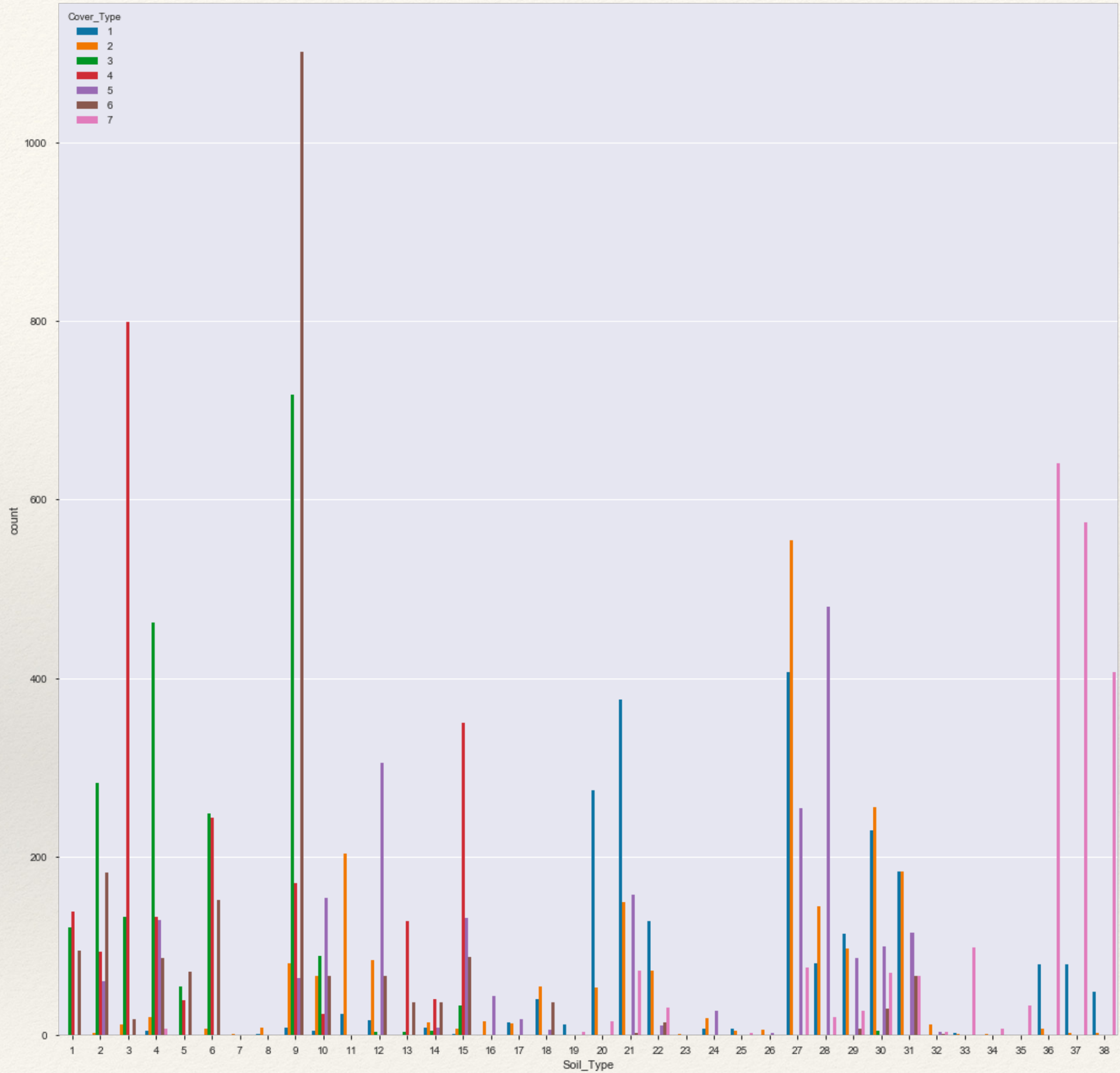
# Distribution





Wilderness Vs CoverType





SoilType Vs CoverType

---

# Conclusion

---

- ❖ Accuracy with **SupportVectorClassifier** : **0.7984457671957672**
- ❖ Accuracy with **ExtraTreesClassifier** : **0.8022486772486772**
- ❖ Accuracy with **RandomForestClassifier** : **0.8492972717609543**
  
- ❖ Future Scope: SVM classifier may perform better with compression techniques like PCA