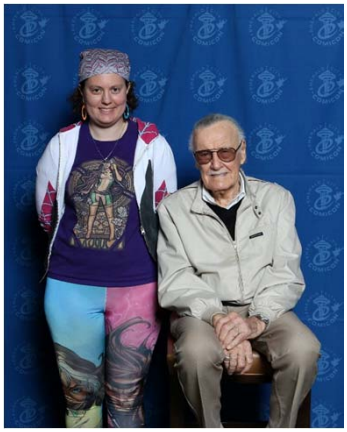
Who Am I?



- 13 years field experience
- Former SharePoint MVP
- Azure Certified
- Former Microsoft
- Code Camp Organizer
- User Group Organizer (Eastside IoT AI ML)

Python

- Released in 1991 by Guido Van Rossum
- Named after Monty Python's Flying Circus
- Very similar to most Object Oriented Languages like C
- Easy for readability and coding
- Good for developers
- More object oriented

R

- Released in 1995 by Ross Ihaka and Robert Gentleman
- Implementation of S programming language by Bell Labs
- Focuses on data analysis, statistical, and graphical modeling
- Mostly used by academics and data scientists
- More functional

Jupyter Notebook

- Python based notebook system to share R, Julia, and Python Projects
- Install using Anaconda
- Prerequisite Python 2.7 or 3.6

R Studio

- Program for Academics
- Free Version
- Can create R Script Files
- Easy to install packages using CRAN

Running Jupyter Notebook/R Studio

- Creating a Jupyter Notebook
- Running R Studio

Importing a CSV File

- R:
  ```
  nba <- read.csv("c:/presentation/nba_2013.csv")
  ```
- Python:
  ```
  import pandas
  nba =
  pandas.read_csv("c:/presentation/nba_2013.csv"
  )
  ```

Finding the Number of Rows

- R:
  dim(nba)
- Python:
  nba.shape

First Row of Data

- R:
  head(nba, 1)
- Python:
  nba.head(1)

Statistical Average

- R:
  sapply(nba, mean, na.rm=TRUE)
- Python:
  nba.mean()

Pairwise Scatterplots

- R:
  library(GGally)
  ggpairs(nba[,c("ast", "fg", "trb")])
- Python:
  import seaborn as sns
  import matplotlib.pyplot as plt
  sns.pairplot(nba[["ast", "fg", "trb"]])
  plt.show()

Clusters of Players

- R:

```
library(cluster) set.seed(1)
isGoodCol <- function(col){
sum(is.na(col)) == 0 && is.numeric(col)
}
goodCols <- sapply(nba, isGoodCol)
clusters <- kmeans(nba[,goodCols], centers=5)
labels <- clusters
$cluster
```

- Python:

```
from sklearn.cluster import KMeans
kmeans_model = KMeans(n_clusters=5, random_state=1)
good_columns = nba._get_numeric_data().dropna(axis=1)
kmeans_model.fit(good_columns)
labels = kmeans_model.labels_
```

Plot Players by Cluster

- R:

```
library(cluster)
nba2d <- prcomp(nba[,goodCols], center=TRUE)
twoColumns <- nba2d$x[,1:2]
clusplot(twoColumns, labels)
```

- Python:

```
import matplotlib.pyplot as plt
from sklearn.decomposition import PCA
pca_2 = PCA(2)
plot_columns = pca_2.fit_transform(good_columns)
plt.scatter(x=plot_columns[:,0], y=plot_columns[:,1], c=labels)
plt.show()
```

Training and Testing Sets

- R:
```
trainRowCount <- floor(0.8 * nrow(nba))
set.seed(1)
trainIndex <- sample(1:nrow(nba), trainRowCount)
train <- nba[trainIndex,]
test <- nba[-trainIndex,]
```
- Python:
```
train = nba.sample(frac=0.8, random_state=1) test =
nba.loc[~nba.index.isin(train.index)]
```

Unvariate Linear Regression

- R:
```
fit <- lm(ast ~ fg, data=train)
predictions <- predict(fit, test)
```
- Python:
```
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(train[["fg"]], train["ast"])
predictions = lr.predict(test[["fg"]])
```

Calculate Summary Statistics for Model

- R:
  ```
  summary(fit)
  ```
- Python:
  ```
  import statsmodels.formula.api as sm
  model = sm.ols(formula='ast ~ fga', data=train)
  fitted = model.fit()
  fitted.summary()
  ```

Python vs R Conclusion

- Python relies mostly on pandas, scikit-learn, and matplotlib
- R has more statistical analysis built-in
- Python has some main packages for data analysis
- R has a lot of smaller packages
- More people are learning Python at 3:1 against R
- Python libraries are more well maintained
- R looks more like something from a statisticians toolkit
- Focusing on Python, because more of a dev
- Might use R for fun with math

All Opinion

- Python is easier to read
- Jupyter Notebook easy to use
- Useable for more than just data science (web development too)
- Object oriented and looks like more common languages
- More courses available

Questions

- ???

References

- NBA Sample Website: https://www.dataquest.io/blog/python-vs-r/
- Datacamp R vs Python: https://www.datacamp.com/community/tutorials/r-or-python-for-data-analysis
- Jupyter Notebooks: http://jupyter.org/
- R Studio: https://www.rstudio.com/
- Data Camp: http://www.datacamp.com

Contact Info

- Twitter: @undiscovereddev
- Personal Email: rebecca.Isserman@gmail.com
- Work Email: rebecca.Isserman@teamintergen.com
- Company Site: http://www.intergen.co.nz