TCDSB #DataDunkers Students Workshop, November 2023

Peter Beens

Twitter/X: @pbeens
Email: pbeens@gmail.com
Mastodon: mstdn.ca/@pbeens

bit.ly/dd-slides

The original of these slides can be found at https://bit.ly/dd-slides.

Our Goal?

To get you to confidently use Jupyter Notebooks and the Python programming language to analyze data from Pascal Siakam's basketball career with the Toronto Raptors.

Overview

- [Activity Demos](#)
- [Introduction to Jupyter Notebooks](#) (Cell Types, Markdown, Running Python Code)
- [Shoot Some Hoops!](#)
- [How Do We Get the Data?](#) (Internal, CSV, Excel, Webpages, Google Sheets)
- [Activity: Mini Basketball Labyrinth](#)
- [Previewing the Data](#) (Columns, Filtering, Sorting)
- [Visualizing the Data](#) (Bar Charts, Scatter Plots, Pie Charts, Histograms)
- [Basic Statistics](#) (min, max, median, mean)
- [Activity: Mini Hoops Graph](#)

Activity Demos (1/2)

1. [Mini Hoops Activity](#) (Intro Only)
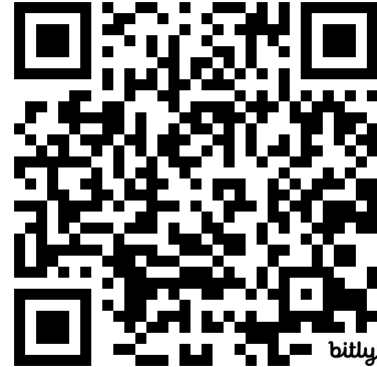2. Mini Basketball Labyrinth (Intro Only)

bit.ly/hoops-data

Activity Demos (2/2)

1. Mini Hoops Activity (Intro Only)
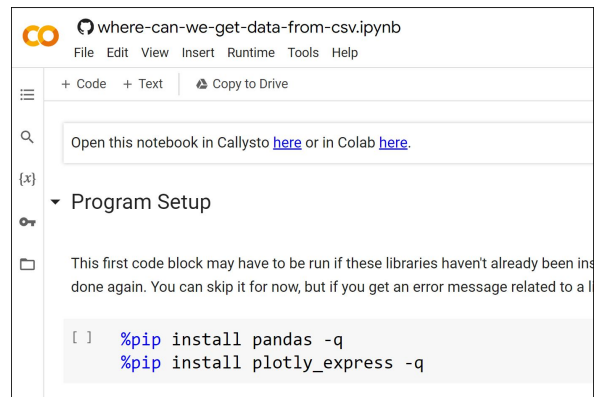2. Mini Basketball Labyrinth (Intro Only)

bit.ly/dd-mini-bb

Key: WNBA

Introduction to Jupyter Notebooks (1/2)

1. [Lesson: Jupyter Notebook Demo](#)
   a. [Callysto](#) vs [Google Colab](#) Discussion
   b. Basics of Markdown
      i. Code Cells vs Markdown Cells *(Pay attention to "Hello world" example program!)*
2. Class Activity: hello-world

Introduction to Jupyter Notebooks (2/2)

1. Lesson: Jupyter Notebook Demo
2. Class Activity: hello-world
   a. Basic Python syntax
   b. How to run code in a code cell

Shoot Some Hoops!

Whenever you see this graphic at the end of a lesson, shoot some hoops and enter it on the form. Clicking on the graphic will take you to the form.

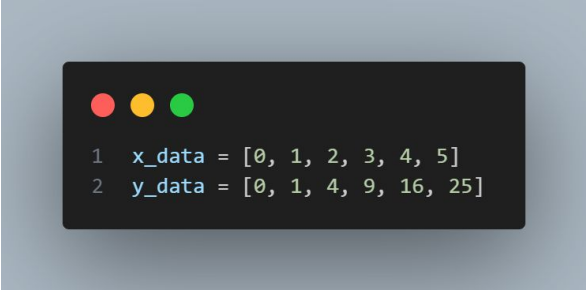Remember to always enter your name the same way, and never enter your last name.

How Do We Get the Data? (1/5)

1. Internal List Data
   a. Shows how to use internal list data ([ ]) to produce a line chart.
2. CSV (Comma Separated Values) File
3. Excel File
4. Webpage
5. Google Sheets

```
1  x_data = [0, 1, 2, 3, 4, 5]
2  y_data = [0, 1, 4, 9, 16, 25]
```

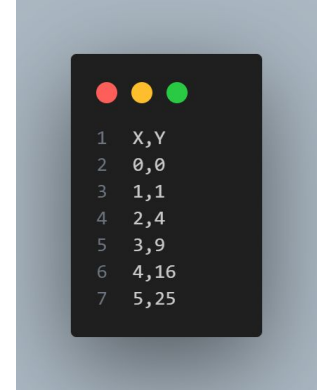# How Do We Get the Data? (2/5)

1. Internal List Data
2. [CSV (Comma Separated Values) File](#)
   a. Shows how to access data from a Comma Separated Values (CSV) file.
   b. Introduces how to use `head()` and `tail()` to show the top or bottom rows, respectively, of your data.
   c. Introduces how to get the name of the columns using `df.columns`.
   d. Introduces how to rename columns.
   e. Introduces Python variables.
3. Excel File
4. Webpage
5. Google Sheets

```
1   X,Y
2   0,0
3   1,1
4   2,4
5   3,9
6   4,16
7   5,25
```

How Do We Get the Data? (3/5)

1. Internal List Data
2. CSV (Comma Separated Values) File
3. Excel File
   a. Shows how to access data from an Excel file.
4. Webpage
5. Google Sheets

	A	B
1	X	Y
2	0	0
3	1	1
4	2	4
5	3	9
6	4	16
7	5	25

How Do We Get the Data? (4/5)

1. Internal List Data
2. CSV (Comma Separated Values) File
3. Excel File
4. Webpage
   a. Shows how to access data from a table on a webpage.
   b. Pay attention to how to choose which table to "scrape".
5. Google Sheets

X	Y
0	0
1	1
2	4
3	9
4	16
5	25

How Do We Get the Data? (5/5)

1. Internal List Data
2. CSV (Comma Separated Values) File
3. Excel File
4. Webpage
5. Google Sheets
   a. Shows how to access data from a Google Sheet.
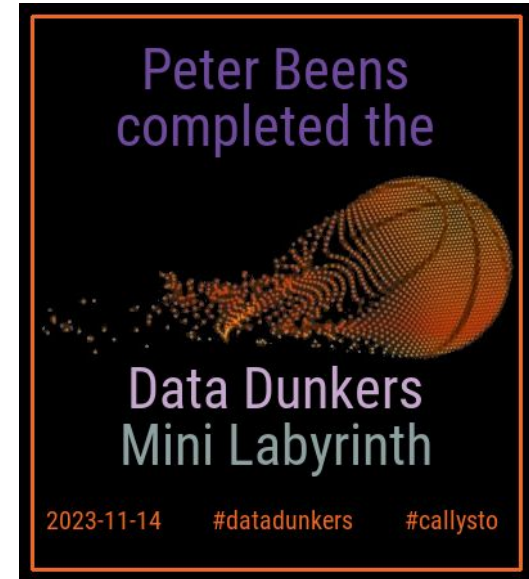   b. Pay attention to the Google Sheet permission and how to change the URL.

Shoot some hoops!

Activity: Mini Basketball Labyrinth

bit.ly/dd-mini-bb

Key: WNBA

Peter Beens
completed the

Data Dunkers
Mini Labyrinth

2023-11-14   #datadunkers   #callysto

Previewing the Data (1/4)

1. **Columns**
   a. How to see what columns are in the data.
   b. How how to view specific columns of the data.
2. Making New Columns
3. Filtering Data
4. Sorting Data

```
1  display(df[['FG', 'FGA']])
```

Previewing the Data (1/4)

1. Columns
2. [Making New Columns](#)
   a. We learn how to make new columns using the data from other columns.
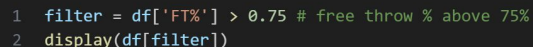   b. We learn how to use `round()` to round numbers.
3. Filtering Data
4. Sorting Data

```
1  # Calculate the shot average for each player and add it as a new column to the data frame
2  raptors_df['Shot Average (%)'] = (raptors_df['FG%'] + raptors_df['FT%']) / 2 * 100
```
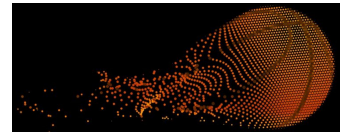
Previewing the Data (3/4)

1. Columns
2. Making New Columns
3. Filtering Data
   a. How to extract data that meets specific criteria.
4. Sorting Data

```
1  filter = df['FT%'] > 0.75 # free throw % above 75%
2  display(df[filter])
```

Previewing the Data (4/4)

1. Columns
2. Making New Columns
3. Filtering Data
4. [Sorting Data](#)
   a. How to sort the data using `df.sort_values()`.
   b. How to delete unwanted rows using `df.drop()`.
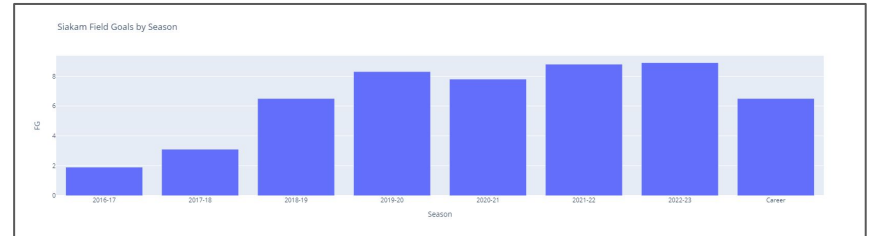
Visualizing the Data (1/4)

1. <u>Bar Charts</u>
   a. How to create a bar chart.
   b. How to plot multiple columns using a list ([ ]).
   c. How to rename the x-axis using `fig.update_xaxes(title='')`.
   d. How to rename the y-axis using `fig.update_yaxes(title='')`.
2. Scatter Plots
3. Pie Charts
4. Histograms

(We have already done Line Charts)
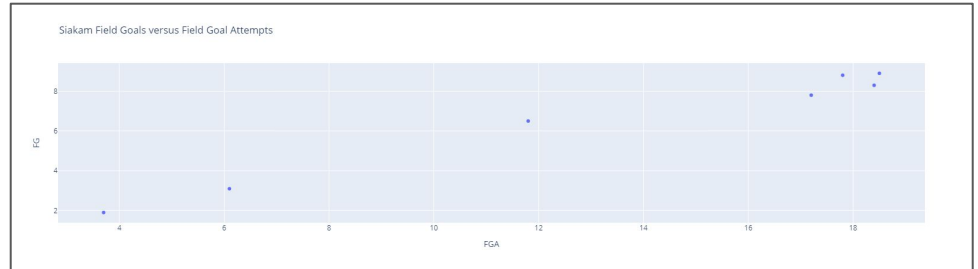
Visualizing the Data (2/4)

1. Bar Charts
2. Scatter Plots
   a. How to create a scatter plot.
   b. How to delete unwanted rows using `df.drop(n, inplace=True)`.
   c. *Method chaining* is introduced.
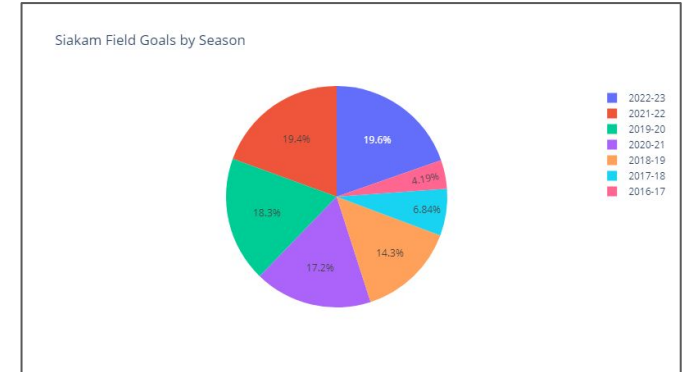   d. The *Plotly Express* `color=` and `size=` options are introduced.
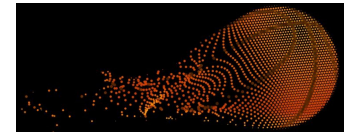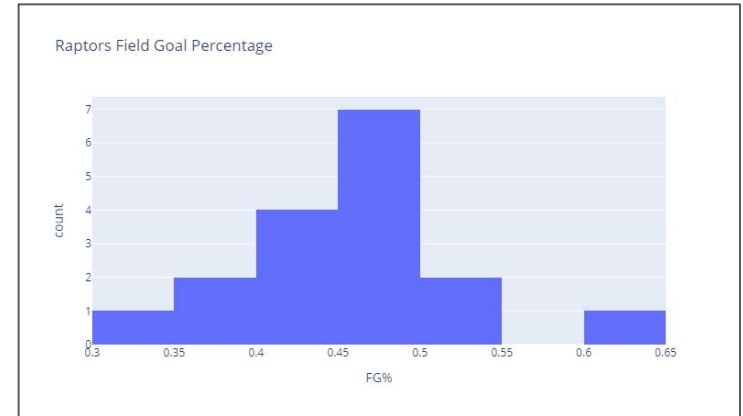3. Pie Charts
4. Histograms

Siakam Field Goals versus Field Goal Attempts

Visualizing the Data (3/4)

Visualizing the Data (4/4)

1. Bar Charts
2. Scatter Plots
3. Pie Charts
4. Histograms
   a. How to create a histogram.
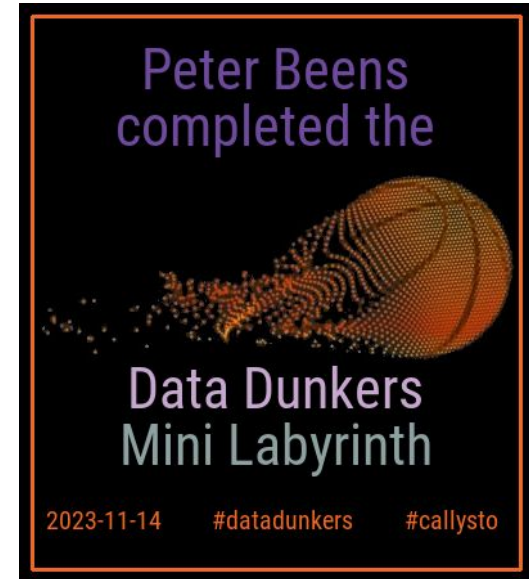

Raptors Field Goal Percentage

Basic Statistics

1. [Basic Statistics](#)
   a. Looking at basic statistics such as `min()`, `max()`, `median()`, `mean()`.
   b. How to look at the stats of all the numeric columns.
   c. How to look at the stats of all columns using `df.describe()`.
   d. How to use *f-strings* to control the number of decimal places that are printed.

Activity: Mini Hoops Graph

bit.ly/hoops-data

Peter Beens completed the

Data Dunkers
Mini Labyrinth

2023-11-14   #datadunkers   #callysto

Thank you!