THINKFUL

2020

Intro to Data Science: The Art of Visualizations

Information and practical exercises to add to your current toolkit or take the first step in launching a new career.

Welcome to Thinkful!

We teach tech skills that lead to fulfilling, high-paying careers.

Our students learn **in-demand** industry tools through **100% online programs** as they work toward a **job-ready portfolio** with the help of an **expert mentor**.

Let's get started.

Workshop Rundown

We're going to talk about:

- ❏ Need for Data Scientists
- ❏ Visualization Principles
- ❏ Relevant Packages in Python
- ❏ Interactive Coding

Big Data By The Numbers

90%	90% of the data in the world today has been created in the last two years alone. [IBM, May 2013]
40K	Google, on average, processes more than 40,000 searches PER SECOND [Forbes, May 2018]
60s	Every minute, we watch 4.1M YouTube videos, send 16M text messages, swipe Tinder 990,000 times, and send $51,892 in transactions on Venmo.

How Is Data Science Useful?

Data Science allows organizations to develop meaningful insights from large amounts of data to help stakeholders make better decisions.

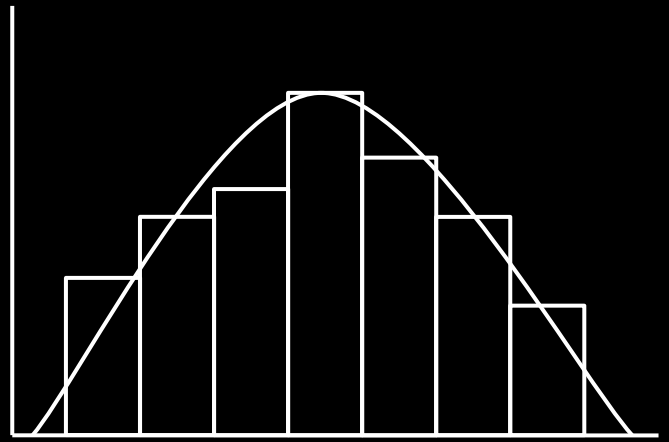
Some specific steps in that process:

- ❑ Data Wrangling
- ❑ Analytics
- ❑ Predictions

Why Are Visualizations Important?

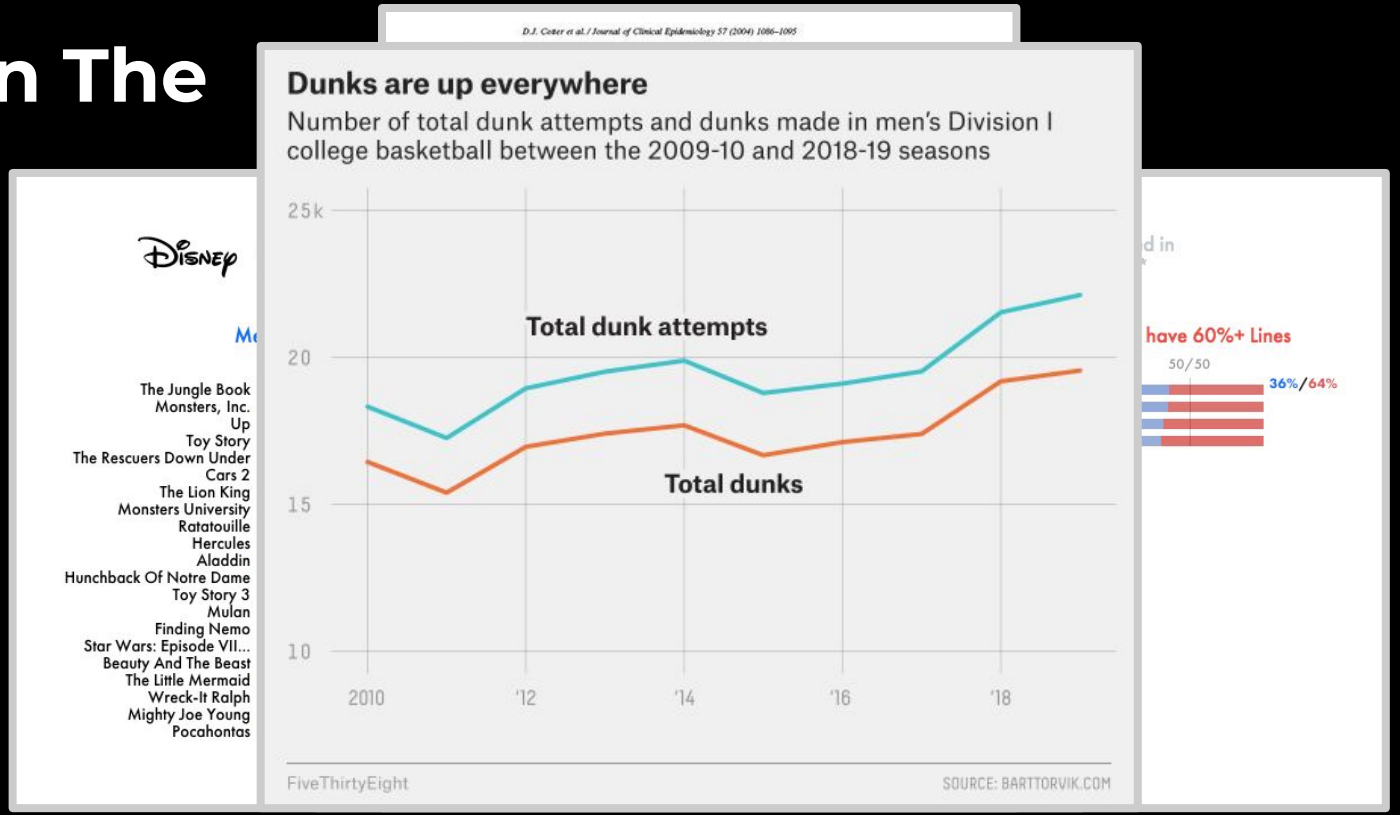
They answer questions about our data:

- For **ourselves**: during data exploration
- For **others**: presenting our work and findings



Decisions In Design

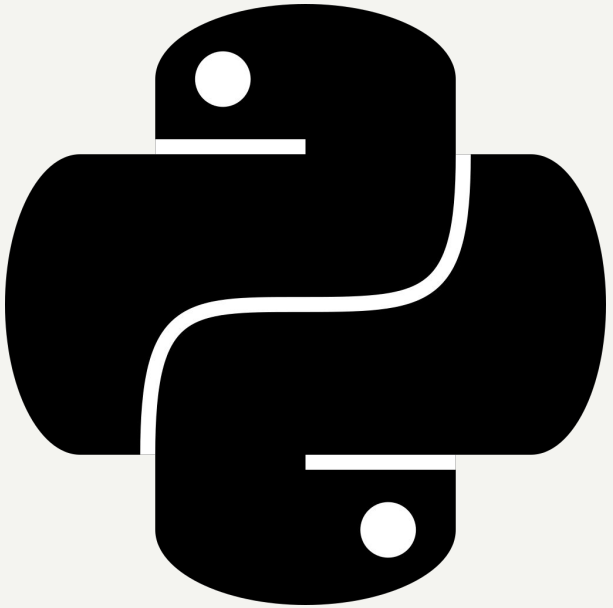
Vizzes In The Wild



Python Basics

Python for Data Science

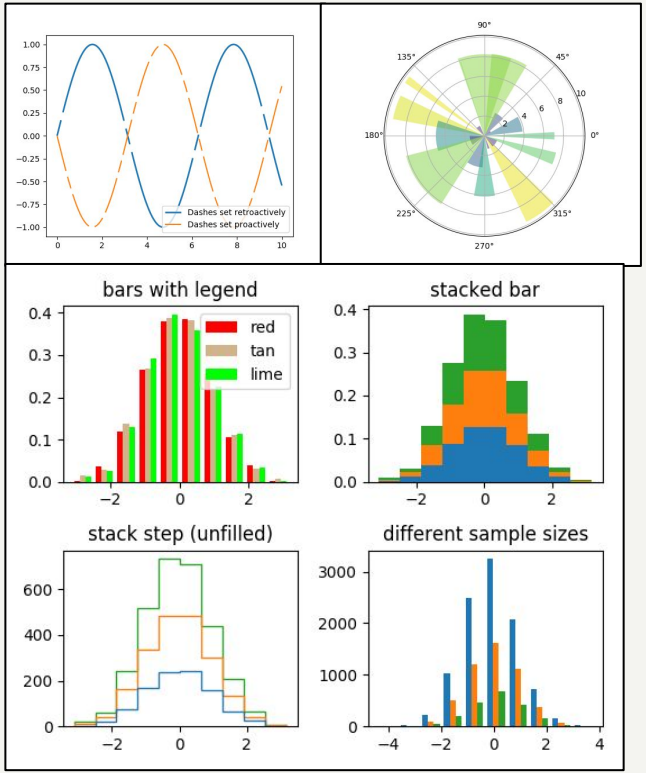
- ❑ Intuitive First Language
- ❑ Customizable Control
- ❑ Robust External Libraries



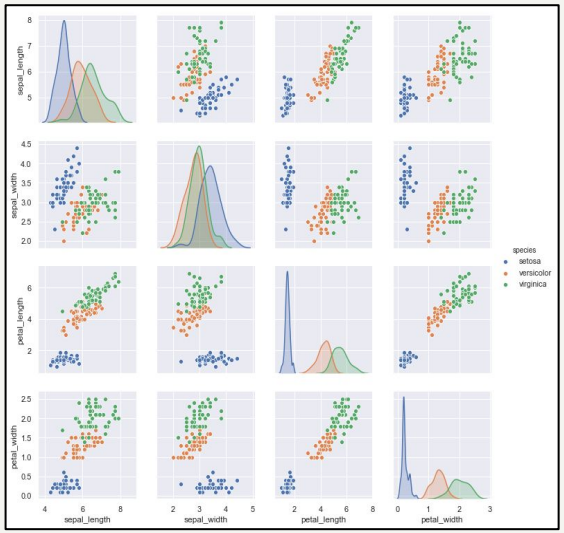
Packages: Matplotlib

Foundational 2D plotting library

- ❑ Makes easy things easy and hard things possible
- ❑ Plots, histograms, power spectra, bar charts, error charts, scatterplots, etc.



Packages: Seaborn



Attractive data visualization library based on Matplotlib

- ❏ High-level interface for drawing attractive and informative statistical graphics.
- ❏ Closely integrated with pandas data structures.
- ❏ Convenient views onto the overall structure of complex datasets

Starter Code Slide

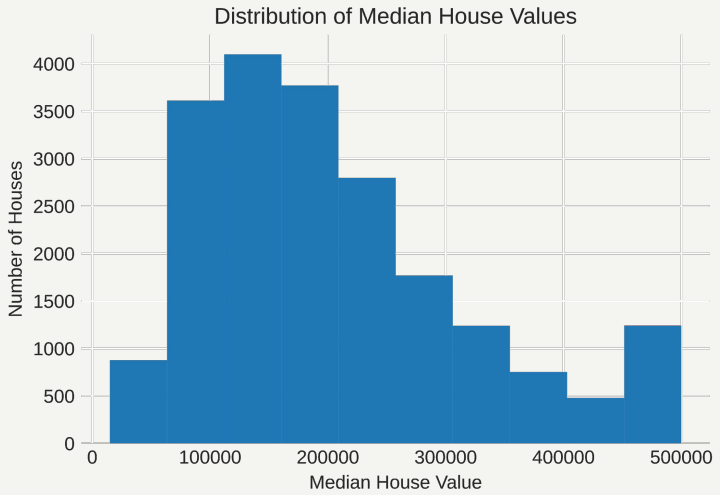
Starter Code
bit.ly/colab_art_of_vis

We'll be using a Google-hosted Python notebook called Colaboratory

- ❑ Click **File**
- ❑ Select **Save a Copy in Drive**
- ❑ This is your personal version of the notebook – let's get started!

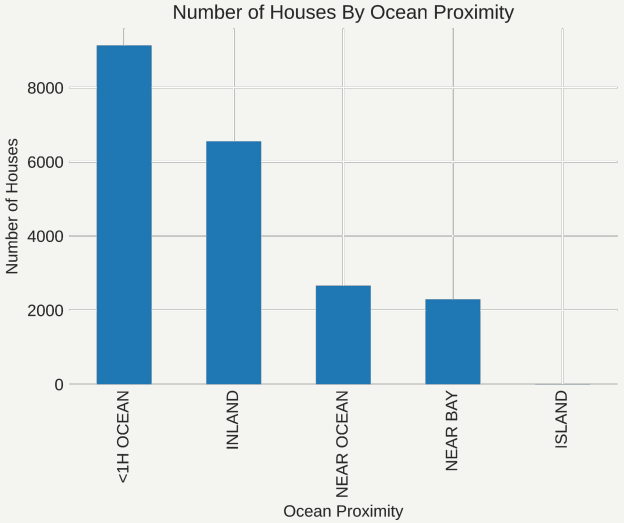
What Is The Distribution Of House Values?

```
# First we create our plot
plt.hist(housing['median_house_value'])

# We can add labels
plt.title('Distribution of Median House Values')
plt.xlabel('Median House Value')
plt.ylabel('Number of Houses')
```



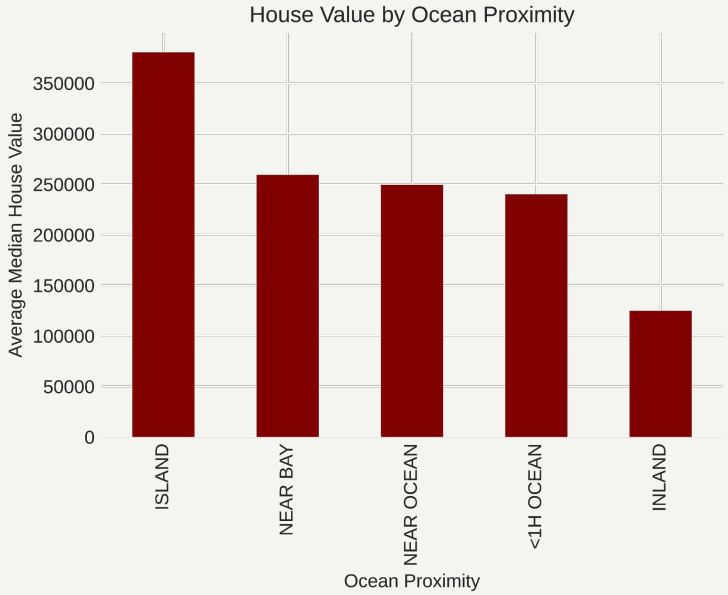
How Close To The Ocean Are The Houses?

```
# Bar chart for those values here:
housing['ocean_proximity'].value_counts().plot(kind = 'bar')

# Labels
plt.title('Number of Houses By Ocean Proximity')
plt.xlabel('Ocean Proximity')
plt.ylabel('Number of Houses');
```



How Does Ocean Proximity Relate To House Value?

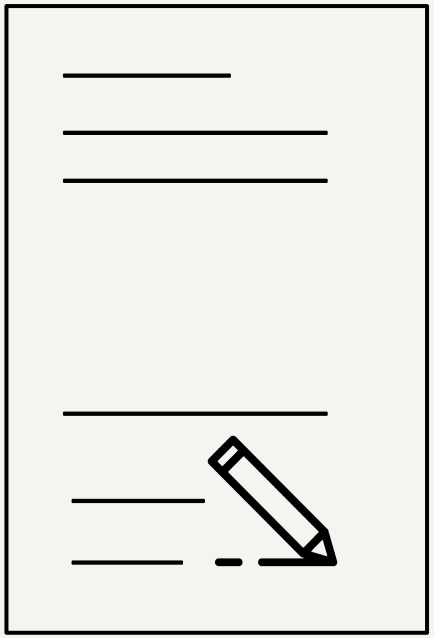
```
# Plotting those values here with bar chart:
ocean_prox_house_val.plot(
    kind = 'bar', color = 'maroon')

# Labels
plt.title('House Value by Ocean Proximity')
plt.xlabel('Ocean Proximity')
plt.ylabel('Average Median House Value');
```



Challenge #1

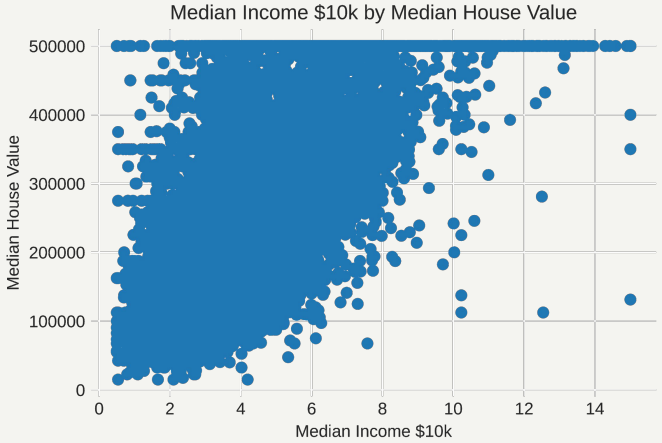
Flex your visualization knowledge:

- ❏ How is another continuous variable related to ocean_proximity?
- ❏ Pick a different column and generate another groupby bar chart with it.
- ❏ Don't forget to update any relevant labels!



Is There A Relationship Between Income And House Value?

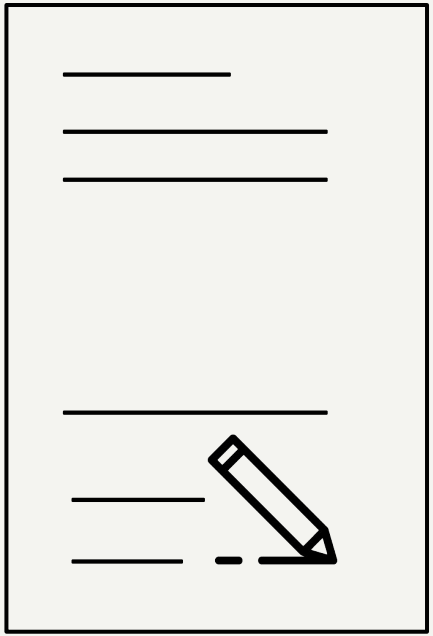
```
# Median Income vs House Value scatter plot
plt.scatter(housing['median_income'], housing['median_house_value'])
plt.title('Median Income $10k by Median House Value')
plt.xlabel('Median Income $10k')
plt.ylabel('Median House Value');
```



Challenge #2

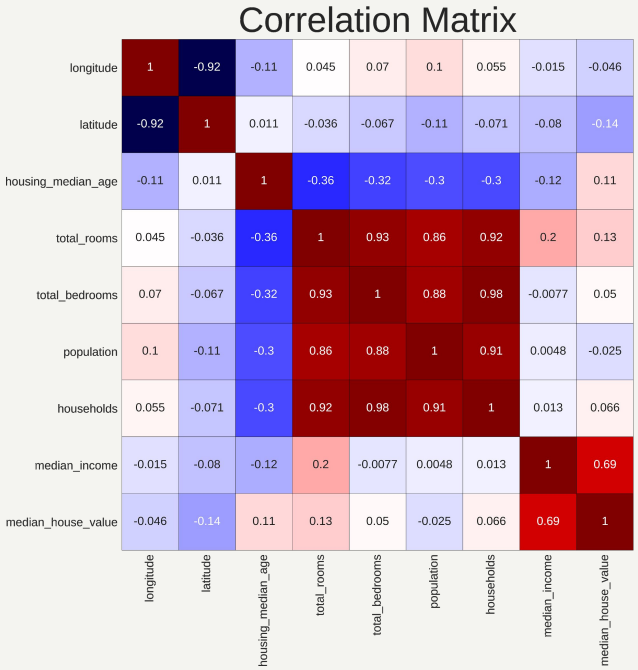
Test your knowledge again:

- ❑ What's another relationship between two continuous variables?
- ❑ Generate another scatter plot to illustrate it.
- ❑ Don't forget to update any relevant labels!



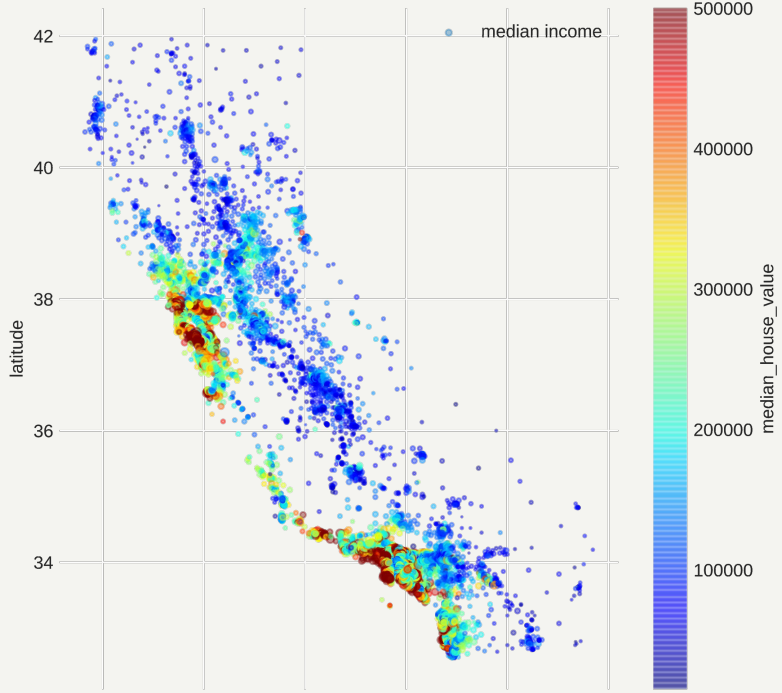
What are the Correlations Between All Continuous Variables?

```
# Visualize our correlation matrix with a heatmap:
plt.figure(figsize=(8,8))
sns.heatmap(housing.corr(),
    linewidths = 0.25,
    square = True,
    linecolor= 'black',
    annot = True);
plt.title('Correlation Matrix', fontsize = 30)
```



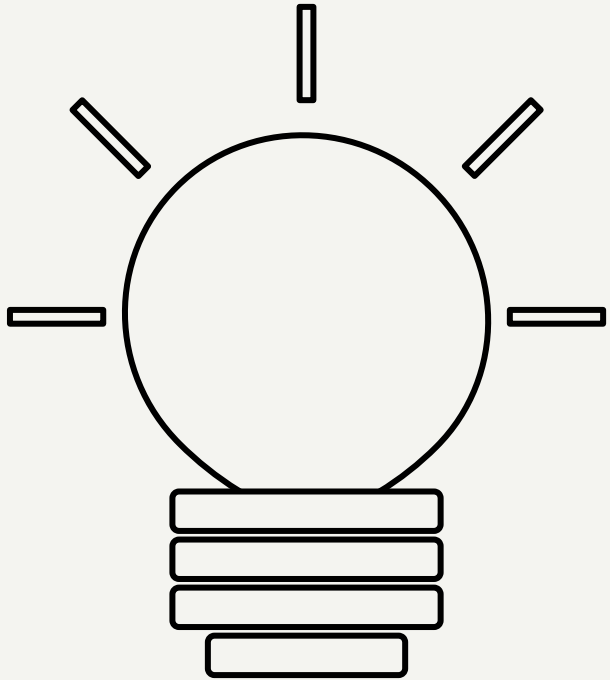
Final Visualization

```
# Creating our final visualization
housing.plot(kind = 'scatter',
    x = 'longitude',
    y = 'latitude'
    figsize = (7,7),
    alpha = 0.4,
    s = housing['median_income'] * 1.5,
    label = 'Median Income',
    c = 'median_house_value',
    cmap = plt.get_cmap('jet'),
    colorbar = True);
plt.title('Correlation Matrix')
```



Today We Learned

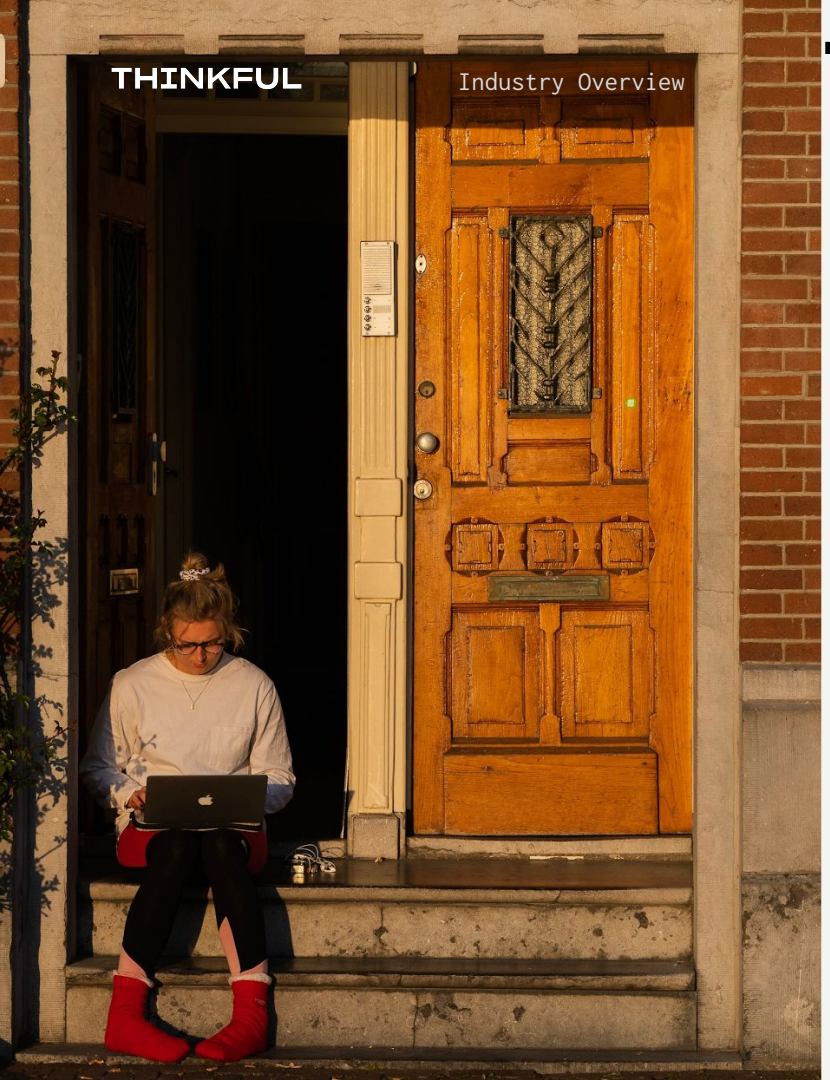
- ❏ Data Science Demand
- ❏ Visualization Principles
- ❏ Histograms
- ❏ Bar Charts
- ❏ Heatmaps
- ❏ Scatter Plots



Common Questions

You might also be wondering

- ❏ What are the outcomes of your students for this field?
- ❏ How do I show my work to a potential employer?
- ❏ Is this course entirely online?
- ❏ What should I do from here?



Take the First Step to A New Career

Anyone who's driven to change their future and achieve a high-earning career is able to enter the world's next workforce. We'll be by your side as you build the skills you need, with personal mentorship and an active, online community of students and educators.

Expand your career opportunities by breaking into tech. Chat with an admissions rep and we'll help you find the perfect fit.

Schedule a Call