Live training:
Introduction to Data Visualization with Python

DataCamp

About me

ADEL NEHME

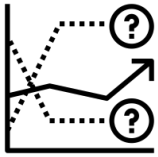
Content Developer

Where does data science bring in value?

Nvidia's StyleGAN2 which generates images of novel faces (link to paper here)

Where does data science bring in value?

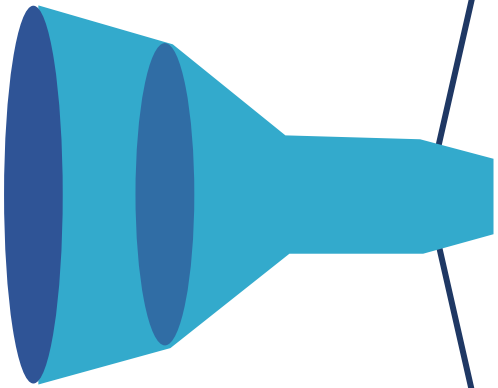
Extend Visibility



Forecast and Act

Data Driven Decision Making

Prioritize what Matters

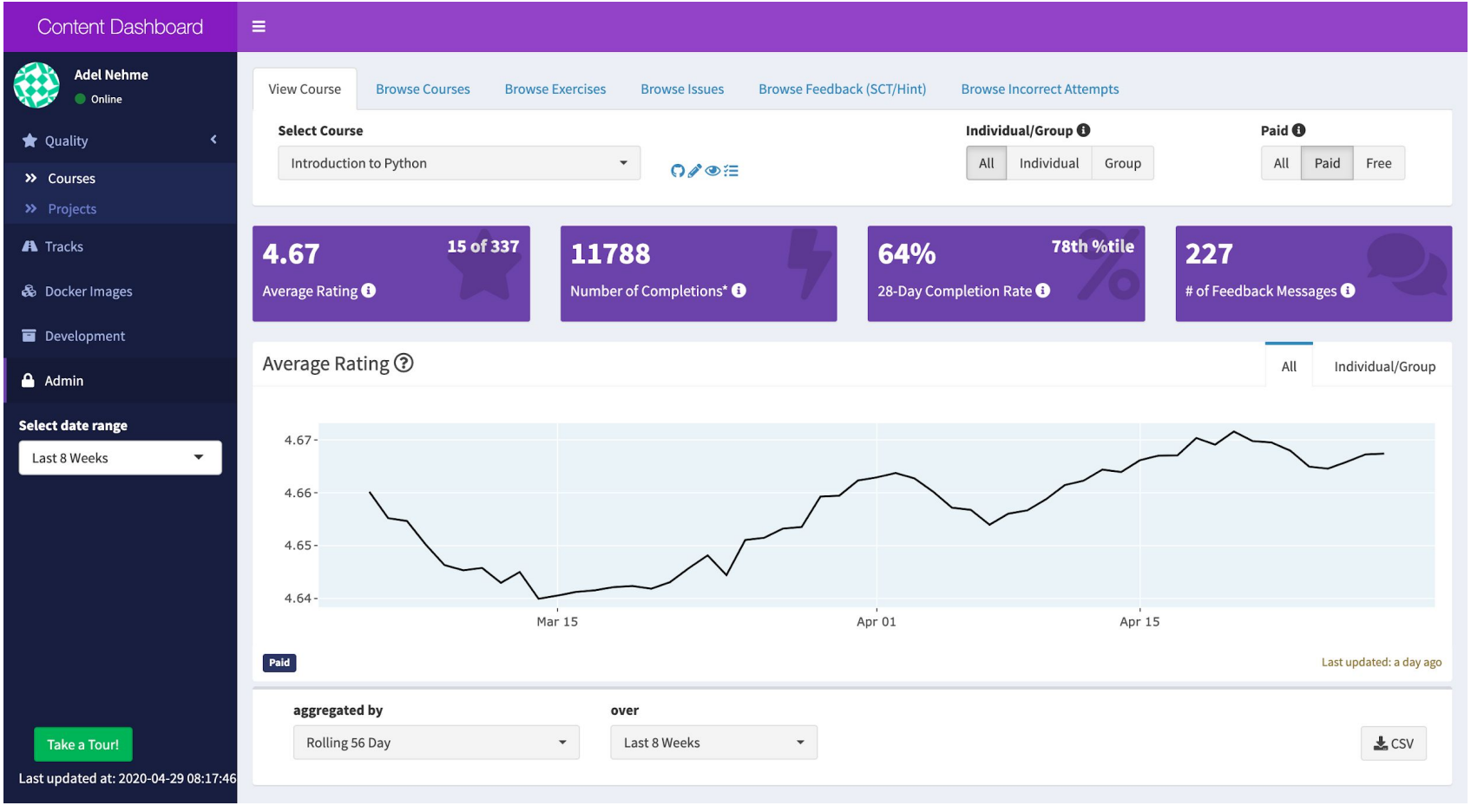


Productivity

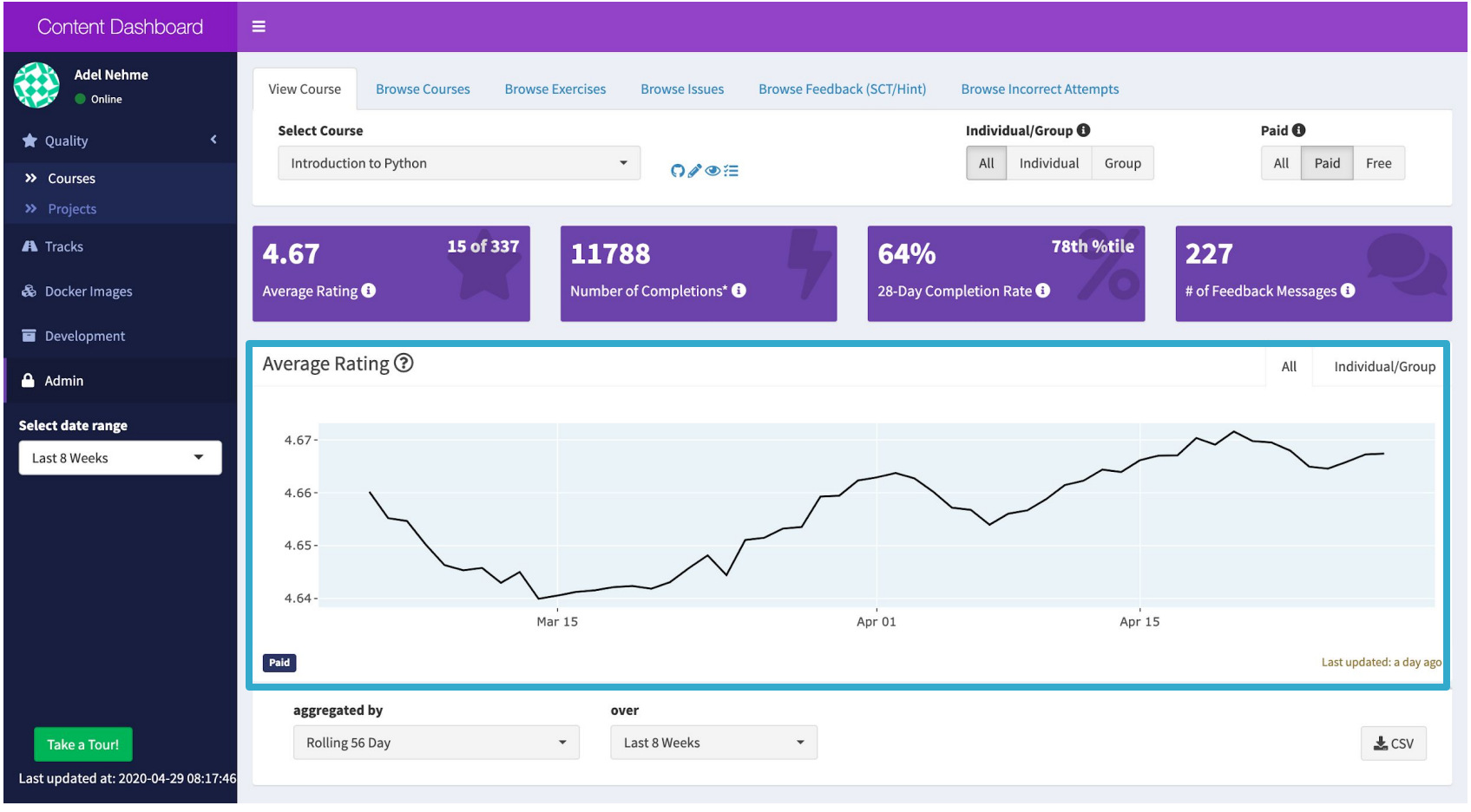
Cut costs

Empowerment

Case in point



Where data visualization comes in



The dataset

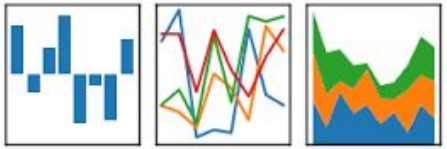
Employee Churn data
Age: *Employee Age*
Attrition: *Stayed or Churned*
DistanceFromHome: *Distance in (km) from home*
Education: *Education level of employee*
EducationField: *Which field did they study*
Gender: *Employee Gender*
MonthlyIncome: *Employee Monthly Income*
PercentSalaryHike: *Percentage of salary increase*
PerformanceRating: *Rating of employee by manager*
YearsAtCompany: *Number of years at company*
EnvironmentSatisfaction: *Survey response on environment satisfaction*
JobInvolvement: *Survey response on job involvement*
JobSatisfaction: *Survey response on job satisfaction*
RelationshipSatisfaction: *Survey response on relationships satisfaction*
WorkLifeBalance: *Survey response on work-life balance*

Human Resources data *about those who left a job vs stayed*

Technologies

pandas

$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



Popular open source data analysis tool for tabular data

matplotlib

Open source plotting library for 2-D visualizations

Seaborn

Open source plotting library built on top of matplotlib

Technologies

jupyter applayout_example Last Checkpoint: 20 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
                      icon='backward',
                      layout=Layout(width='80%',
                                    height='30%'))
next_button = Button(description="Next",
                     icon='forward',
                     layout=Layout(width='80%',
                                   height='30%'))
footer = HTML("Filename: {}".format(image_file))

AppLayout(header=header,
          left_sidebar=prev_button,
          center=image,
          right_sidebar=next_button,
          footer=footer,
          grid_gap='20px',
          justify_items='center',
          align_items='center')
```

Simple Image Viewer

Prev

Next

Filename: images/cat.jpg

jupyter CO

❗❗Requires a gmail account to edit❗❗

Session outline

1. *Introduction*
2. *Some data visualization basics I: The anatomy of a plot*
3. *Distribution plots*
4. *Some data visualization basics II: Subplots*
5. *Multiple Distribution plots*
6. *Q&A*
7. *Data Visualization*
8. *Q&A*
9. *Data Visualization*
10. *Q&A*
11. *Closing notes*

Notebook

Session outline

1. *Introduction*
2. *Some data visualization basics I: The anatomy of a plot*
3. *Distribution plots*
4. *Some data visualization basics II: Subplots*
5. *Multiple Distribution plots*
6. *Q&A*
7. *Data Visualization*
8. *Q&A*
9. *Data Visualization*
10. *Q&A*
11. *Closing notes*

Closing notes

Understand the basics ...

Figure

Figure

Understanding how plots work enables you to ask the right questions

… look up the rest!

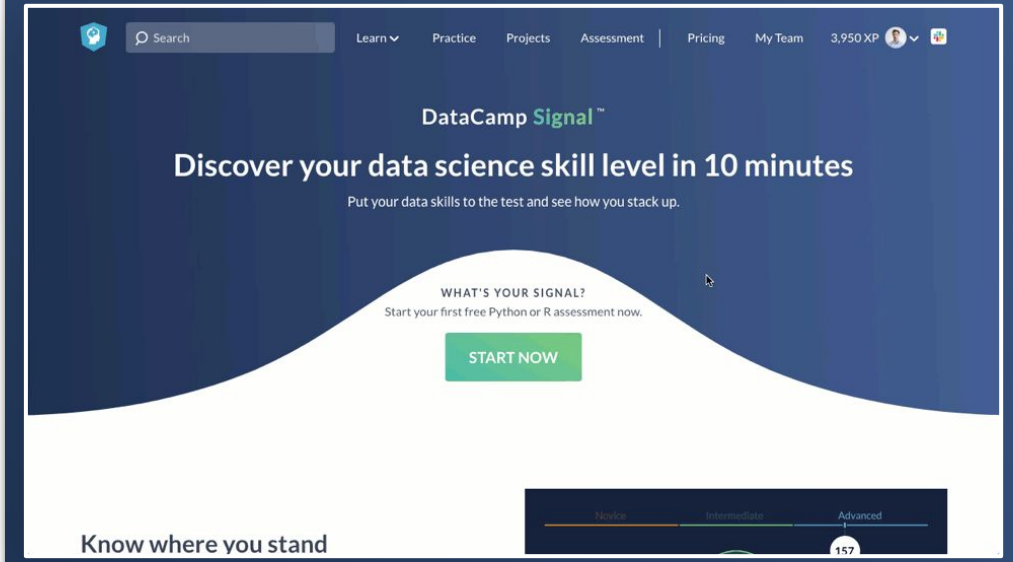
Ask the right questions in the right places:

- Google is your friend
- Stackoverflow
- Blog Posts
- DataCamp slack community
- DataCamp courses!
- Seaborn documentation
- Matplotlib documentation

Discover your team's data skill level with Signal™

An adaptive assessment tool that gives you a bird's-eye view of your team's skill level

- Learners can discover their skill level in just 10 minutes
- Signal is a quick but rigorous assessment that requires learners to write actual code in addition to multiple-choice questions
- The difficulty automatically adjusts based on performance
- Learners can explore detailed results to see how their score was calculated, pinpoint where they made mistakes, and track progress over time
- Signal provides personalized course recommendations based on learners' strengths and skill gaps
- Now in beta: We've launched a new assessment reporting page to help you easily visualize your team's performance, along with recommended courses based on your team's skill gaps



Work with our content team to find courses perfect for your organization

Consumers

- Intro to SQL for Data Science
- Data Analysis with Spreadsheets
- Data Visualization in Spreadsheets

Leaders

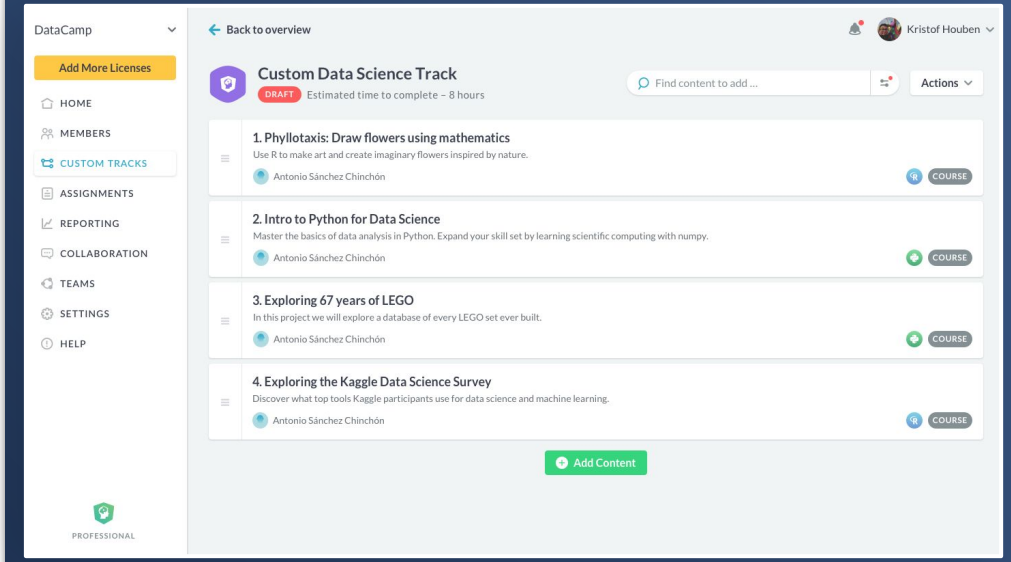
- Data Science for Managers
- Numeral Decision-Making (set to Launch 11/2019)
- Machine Learning for Managers (set to Launch 12/2019)

Analysts

- Career Track: Data Scientist with Python / or "R"
- Supervised learning with scikit-learn
- Machine learning toolbox

Citizen Data Scientists

- Streamline Data ingestion with pandas
- Feature Engineering with PySpark
- Visualizing Big Data with Trelliscope



Content for Business

Remove the bottlenecks to data fluency with our content designed for businesses

- Our **Tableau for Business** training helps companies tell better data stories
- Our **Oracle for Business** training equips your team with the skills to extract relevant insights
- Our **Data Skills for Business** skill track helps decision-makers apply data to solve business problems in a more informed way

Interactive course and learning tracks

Modern learn-by-doing approach

- In-browser coding exercises
- Bite-sized videos with expert instructors
- Real-time AI-powered feedback

325+ unmatched courses to build data fluency

python™ R SQL SPREADSHEETS

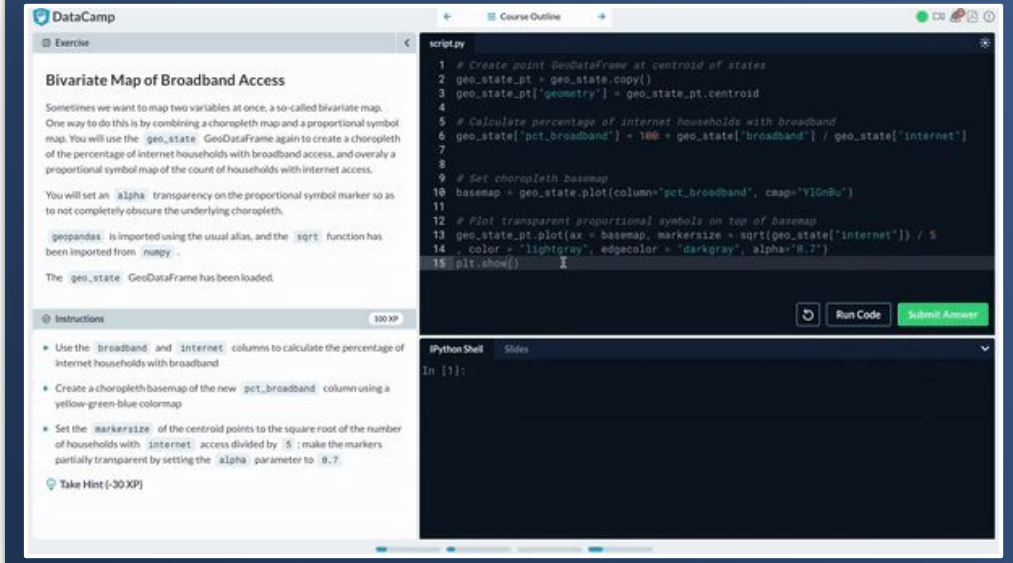
git Shell Scala Spark

50+ skill and career tracks

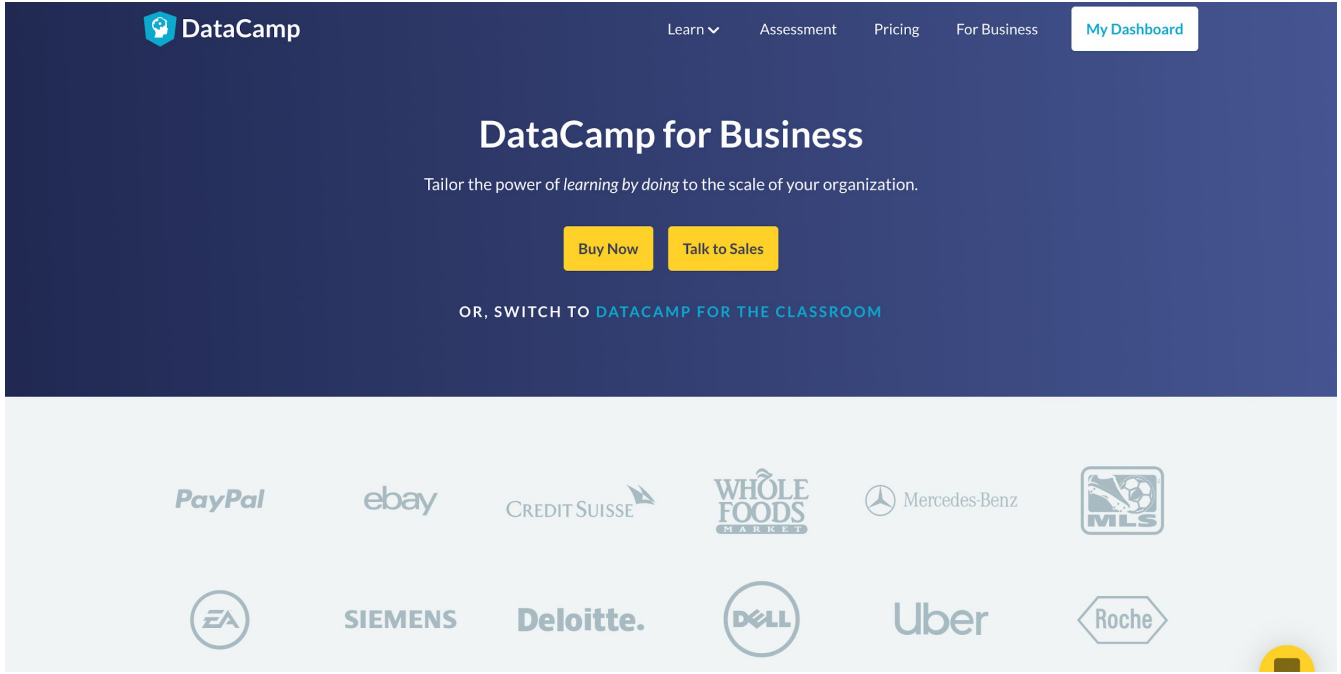
- Curated course collections for guided learning

275+ industry-leading instructors

- Learn from industry experts and authors of renowned code packages



Check out DataCamp's for Business plan



Request a demo here!

Take home question

Here's the link to a fuller dataset (requires a Kaggle account)

1) *What are other potential drivers of Churn? Be creative and visualize your reports!*

Functions that should/could be used:

- Here are some visualizations at your disposal!

***Bonus points** if* you make your visualizations extra pretty!

Submission details:

- Share with us a code snippet with your output on LinkedIn, Twitter or Facebook
- Tag us on @DataCamp or @Adel_Nehme with the hashtag `#datacamplive`

Recap of the functions used

Visualization Functions	Description
sns.distplot()	Creates a distribution plot
sns.boxplot()	Creates a boxplot
sns.swarmplot()	Creates a swarmplot
sns.barplot()	Creates a barplot
sns.scatterplot()	Creates scatterplots
sns.lineplot()	Creates a lineplot

Pandas Functions	Description
pd.crosstab()	Find # of row by values of 2 columns
pd.cut()	Cut a continuous variable into categories

Visualization helper functions	Description
sns.set_style()	Set the style of a visualization
sns.despine()	Remove axis on a plot
figure, axes = plt.subplots()	Creates a figure and arbitrary # of axes
plt.axvline()	Creates a vertical line
plt.text()	Adds text to a visualization
plt.xlabel()	Set custom x-axis label
plt.ylabel()	Set custom y-axis label
plt.title()	Set custom title for plot