Data Visualization

Data Analysis Steps (From Udacity ND)

1. Question
2. Wrangle
3. Explore
4. Draw Conclusions
5. Communicate

Data Analysis Steps (From Udacity ND)

Step 1: Ask questions

Either you're given data and ask questions based on it, or you ask questions first and gather data based on that later. In both cases, great questions help you focus on relevant parts of your data and direct your analysis towards meaningful insights.

Step 2: Wrangle data

You get the data you need in a form you can work with in three steps: gather, assess, clean. You gather the data you need to answer your questions, assess your data to identify any problems in your data's quality or structure, and clean your data by modifying, replacing, or removing data to ensure that your dataset is of the highest quality and as well-structured as possible.

Step 3: Perform EDA (Exploratory Data Analysis)

You explore and then augment your data to maximize the potential of your analyses, visualizations, and models. Exploring involves finding patterns in your data, visualizing relationships in your data, and building intuition about what you're working with. After exploring, you can do things like remove outliers and create better features from your data, also known as feature engineering.

Step 4: Draw conclusions (or even make predictions)

This step is typically approached with machine learning or inferential statistics that are beyond the scope of this course, which will focus on drawing conclusions with descriptive statistics.

More on machine learning: Machine Learning Engineer Nanodegree

Step 5: Communicate your results

You often need to justify and convey meaning in the insights you've found. Or, if your end goal is to build a system, you usually need to share what you've built, explain how you reached design decisions, and report how well it performs. There are many ways to communicate your results: reports, slide decks, blog posts, emails, presentations, or even conversations. Data visualization will always be very valuable.

Before walking through each of these steps with real datasets using Python, let's build a bit of

From Google Data Analytics Certificate

It is time to enter the **data analysis life cycle**—the process of going from data to decision. Data goes through several phases as it gets created, consumed, tested, processed, and reused. With a life cycle model, all key team members can drive success by planning work both up front and at the end of the data analysis process. While the data analysis life cycle is well known among experts, there isn't a single defined structure of those phases. There might not be one single architecture that's uniformly followed by every data analysis expert, but there are some shared fundamentals in every data analysis process. This reading provides an overview of several, starting with the process that forms the foundation of the Google Data Analytics Certificate.

The process presented as part of the Google Data Analytics Certificate is one that will be valuable to you as you keep moving forward in your career:

1. **Ask**: Business Challenge/Objective/Question
2. **Prepare**: Data generation, collection, storage, and data management
3. **Process**: Data cleaning/data integrity
4. **Analyze**: Data exploration, visualization, and analysis
5. **Share**: Communicating and interpreting results
6. **Act**: Putting your insights to work to solve the problem

Understanding this process—and all of the iterations that helped make it popular—will be a big part of guiding your own analysis and your work in this program. Let's go over a few other variations of the data analysis life cycle.

From EMC

EMC's data analysis life cycle

EMC Corporation's data analytics life cycle is cyclical with six steps:

1. Discovery
2. Pre-processing data
3. Model planning
4. Model building
5. Communicate results
6. Operationalize

EMC Corporation is now Dell EMC. This model, created by David Dietrich, reflects the cyclical nature of real-world projects. The phases aren't static milestones; each step connects and leads to the next, and eventually repeats. Ke questions help analysts test whether they have accomplished enough to move forward and ensure that teams hav spent enough time on each of the phases and don't start modeling before the data is ready. It is a little different fr the data analysis life cycle this program is based on, but it has some core ideas in common: the first phase is intere in discovering and asking questions; data has to be prepared before it can be analyzed and used; and then finding should be shared and acted on.

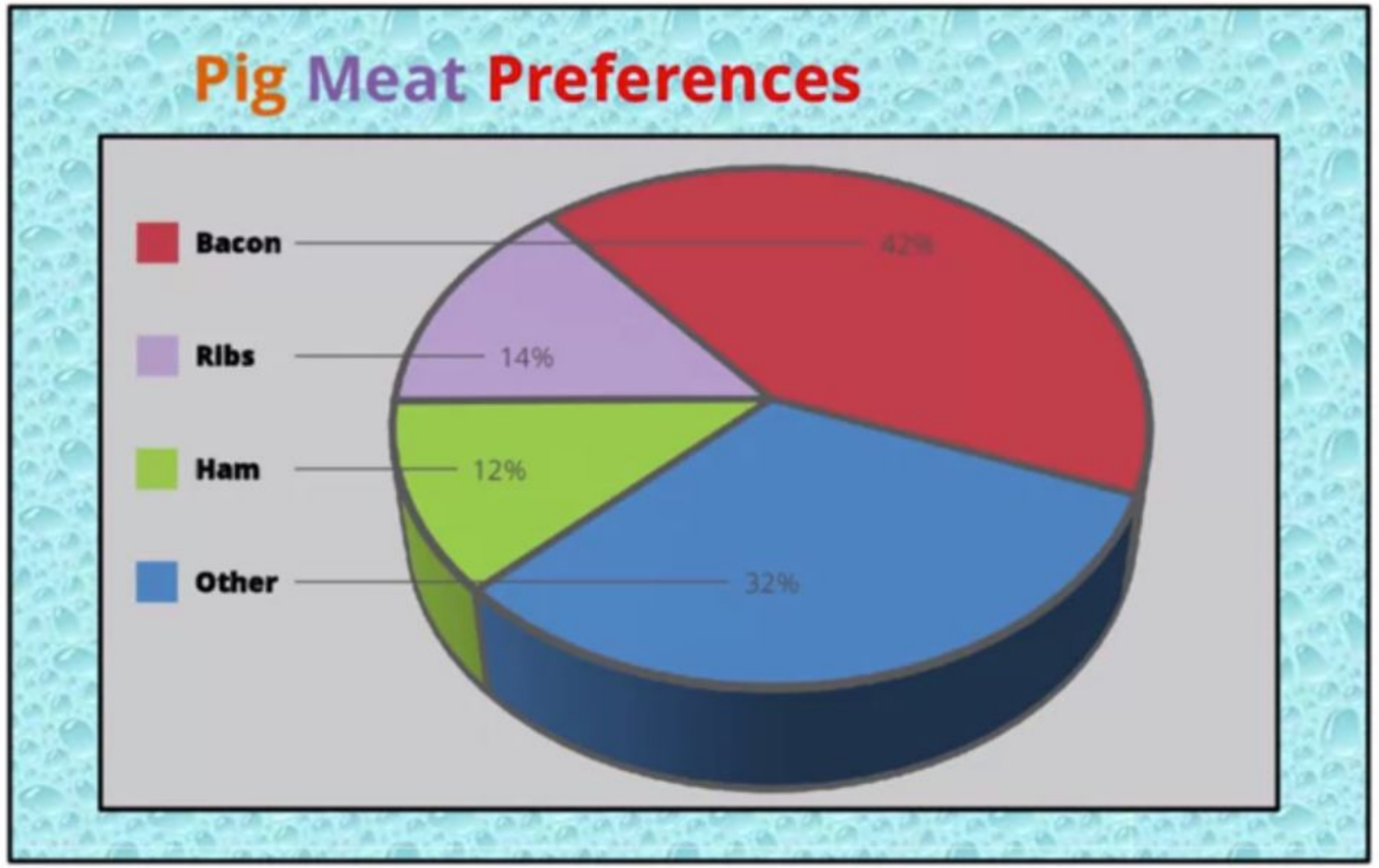
For more information, refer to this e-book, Data Science & Big Data Analytics.

Why build visuals?

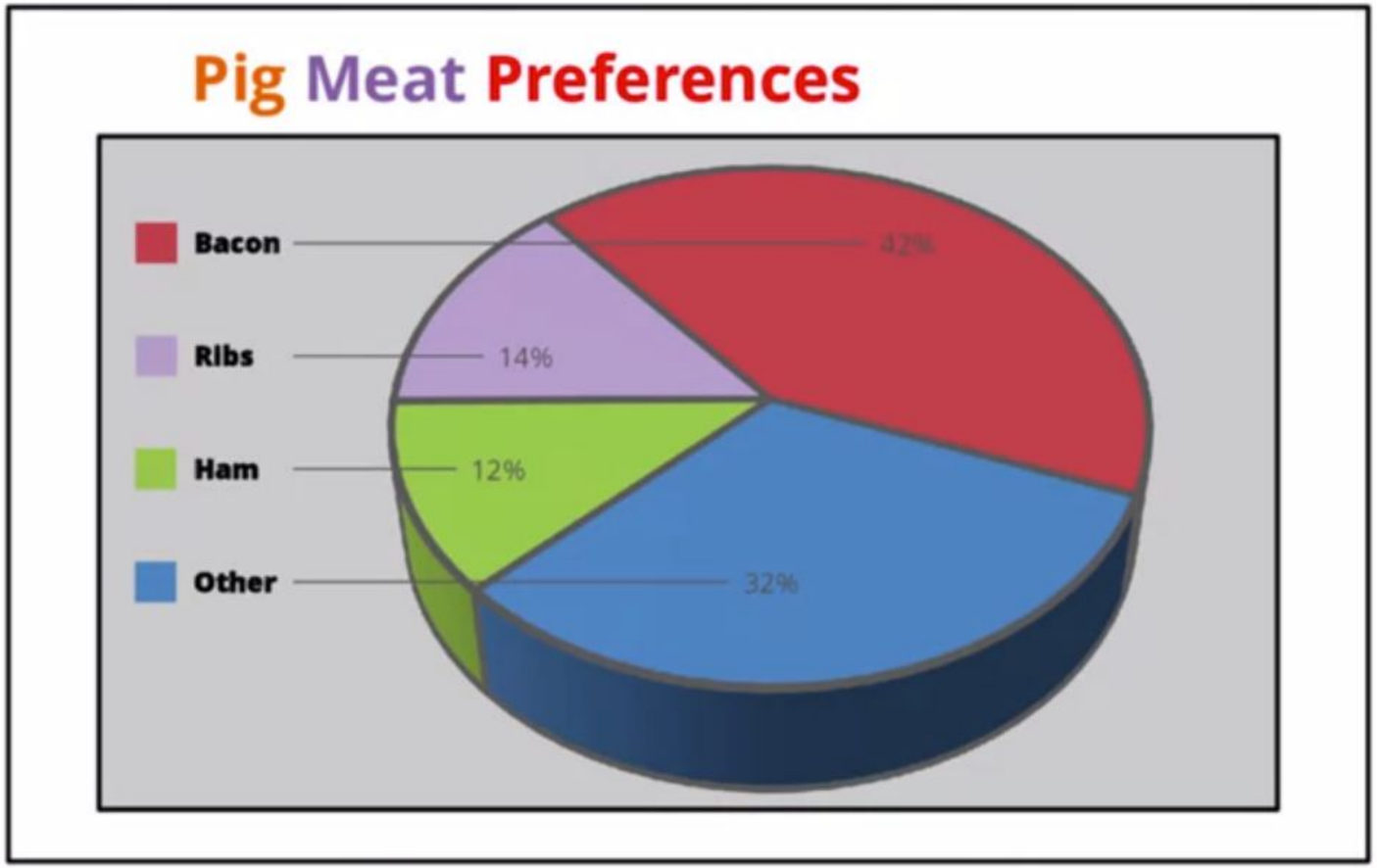
- A picture is worth a thousand words
- For exploratory data analysis
- Communicate data clearly
- Share unbiased representation of data
- Support recommendations to different stakeholders

Visualization Example

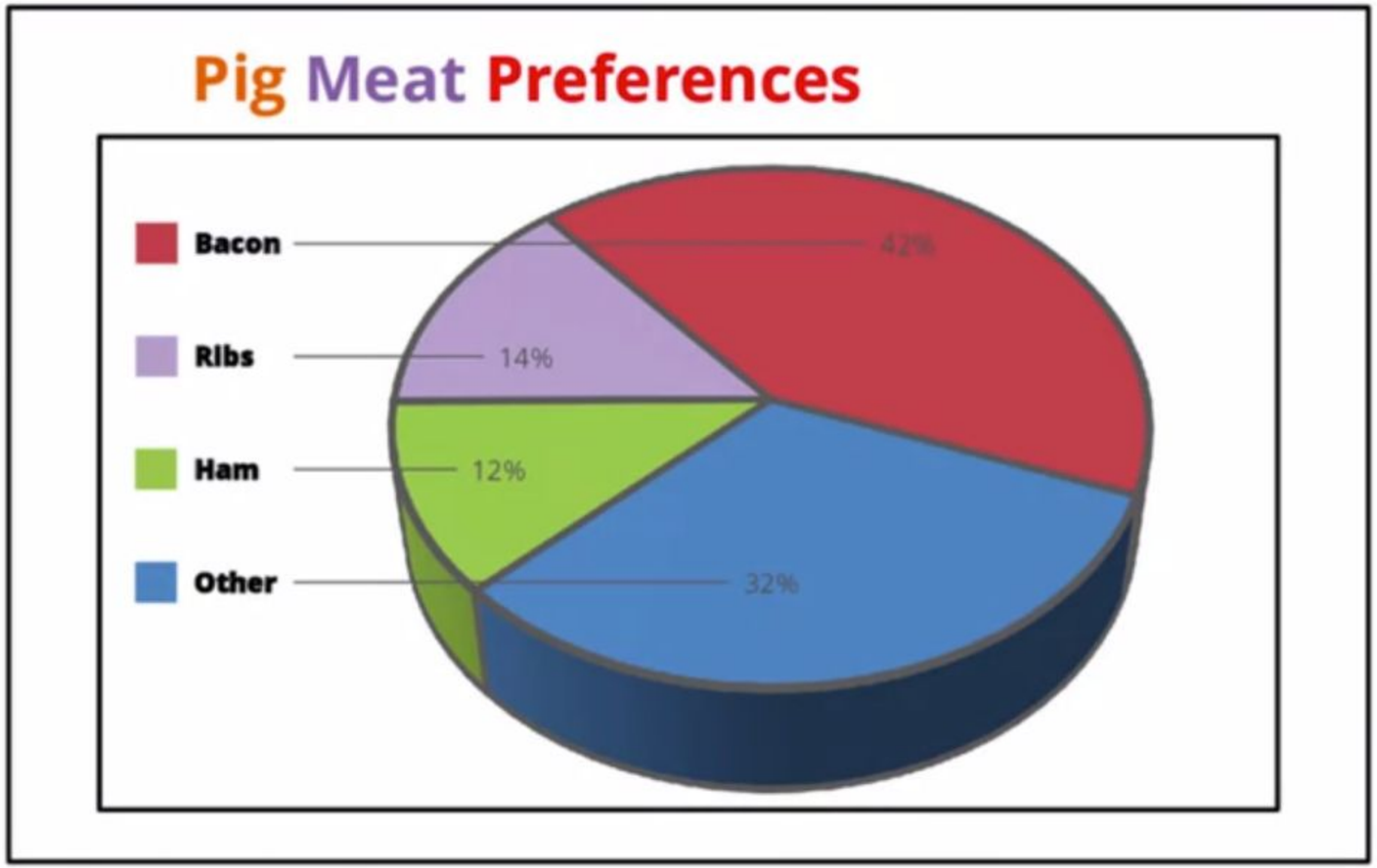
Visualization Example (1/12)



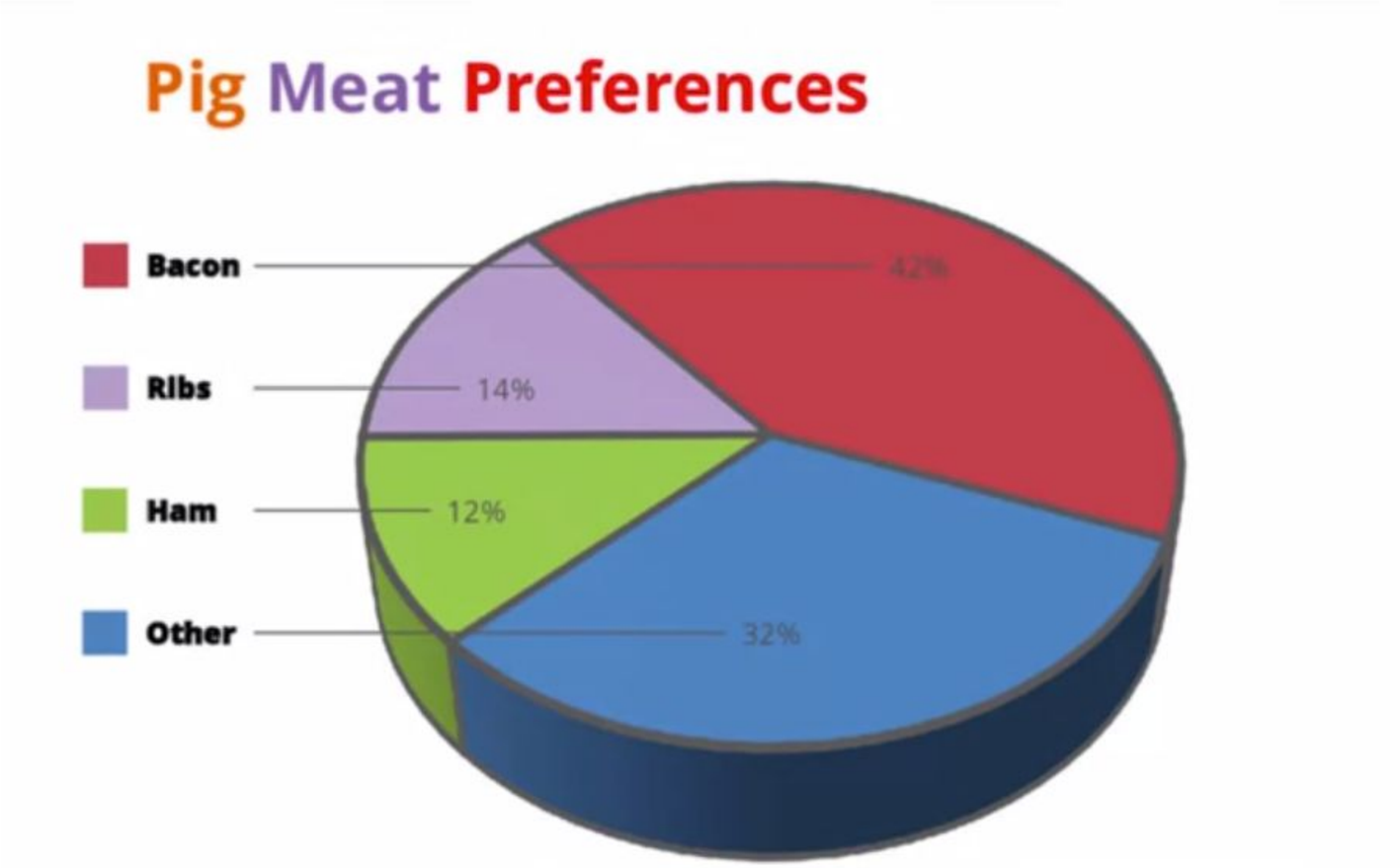
Visualization Example (2/12)



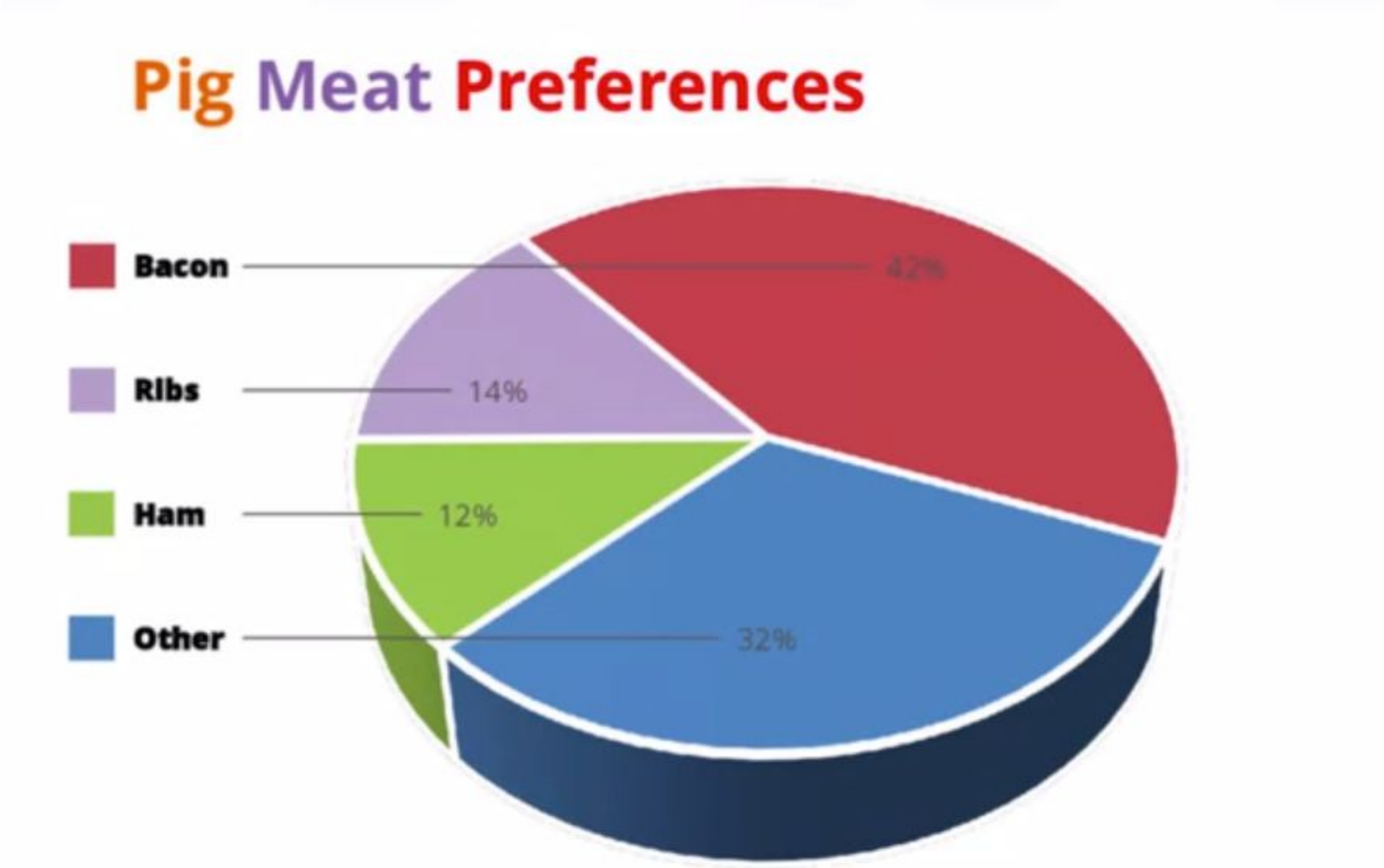
Visualization Example (3/12)



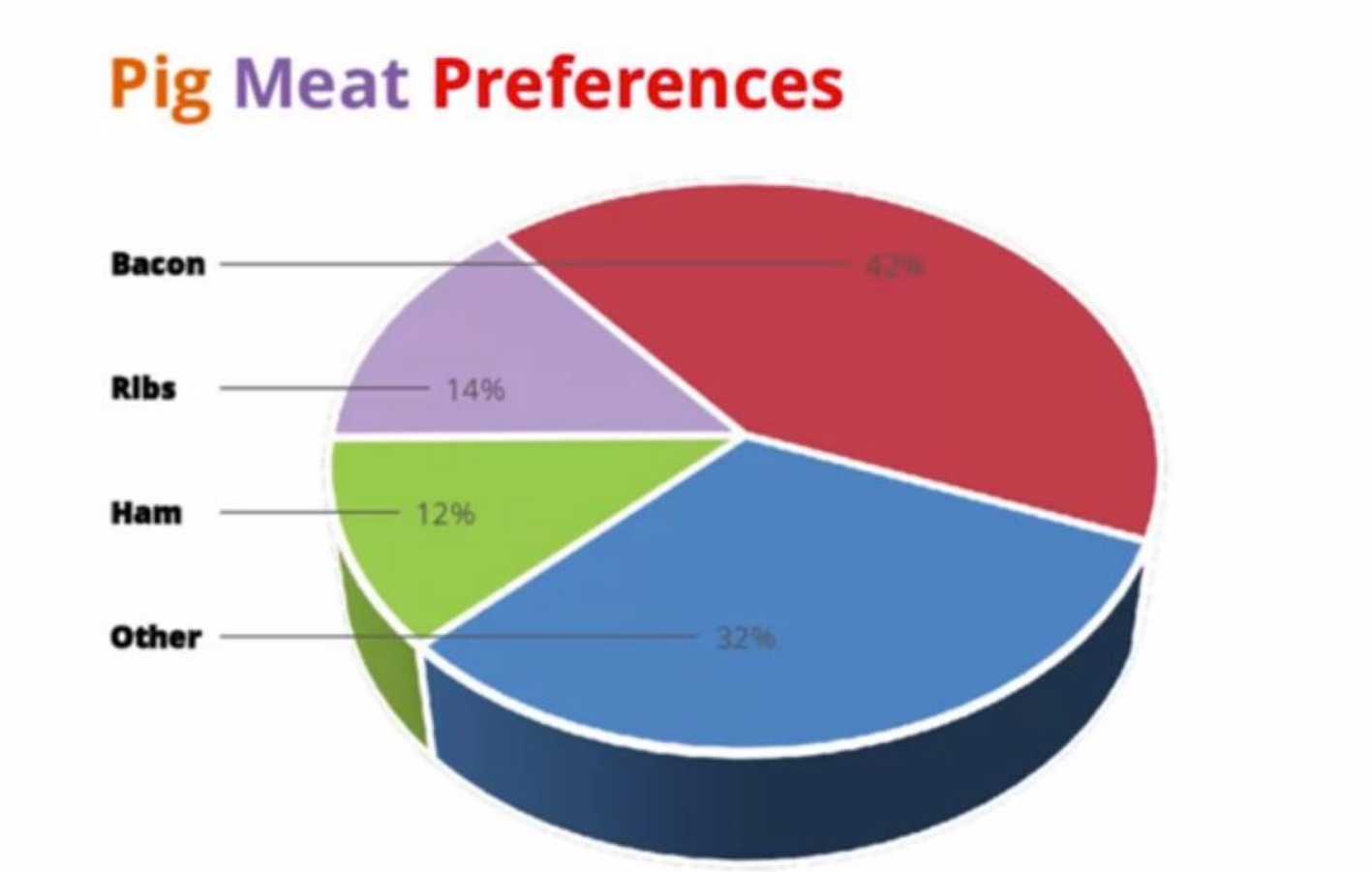
Visualization Example (4/12)



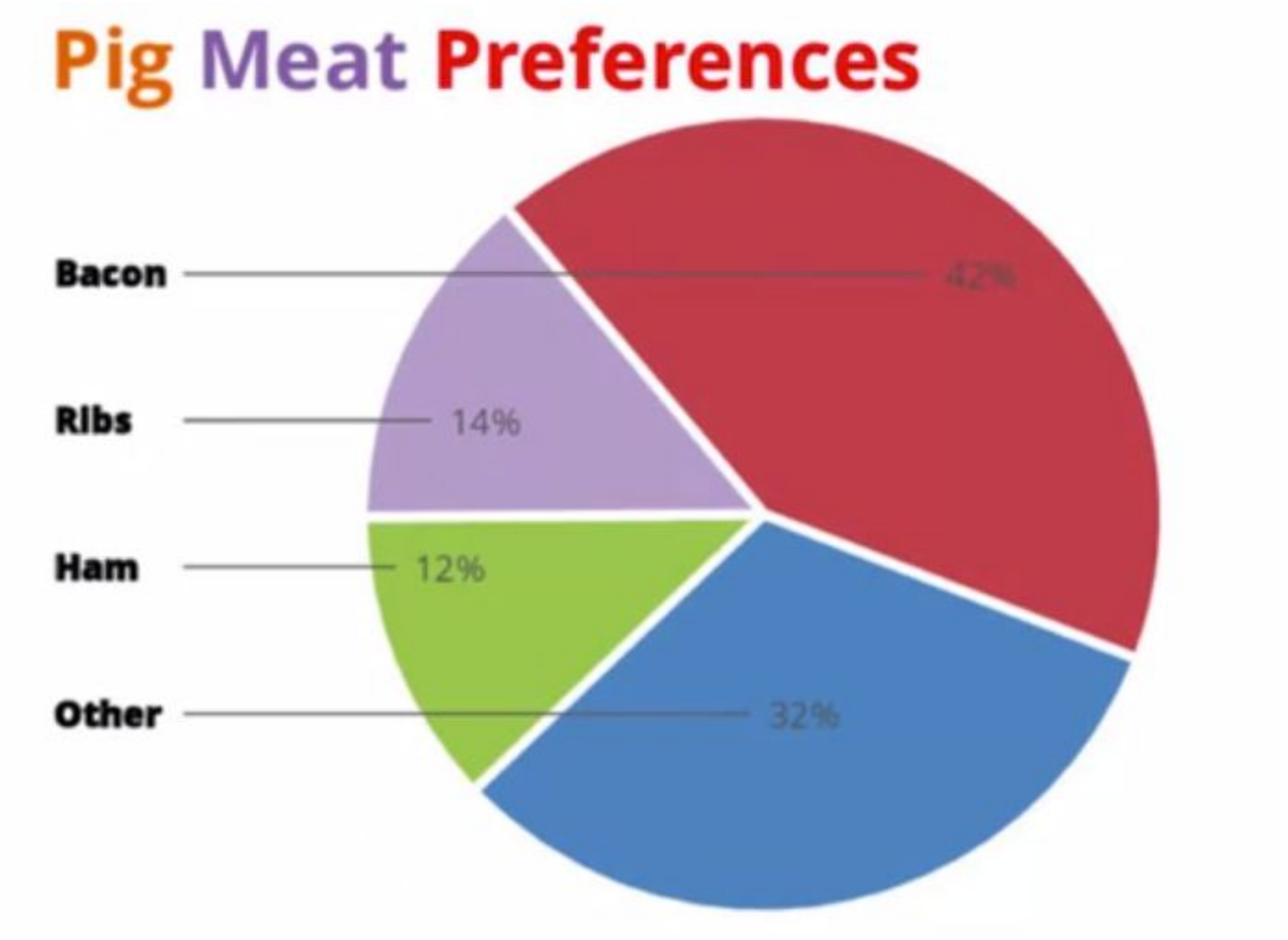
Visualization Example (5/12)



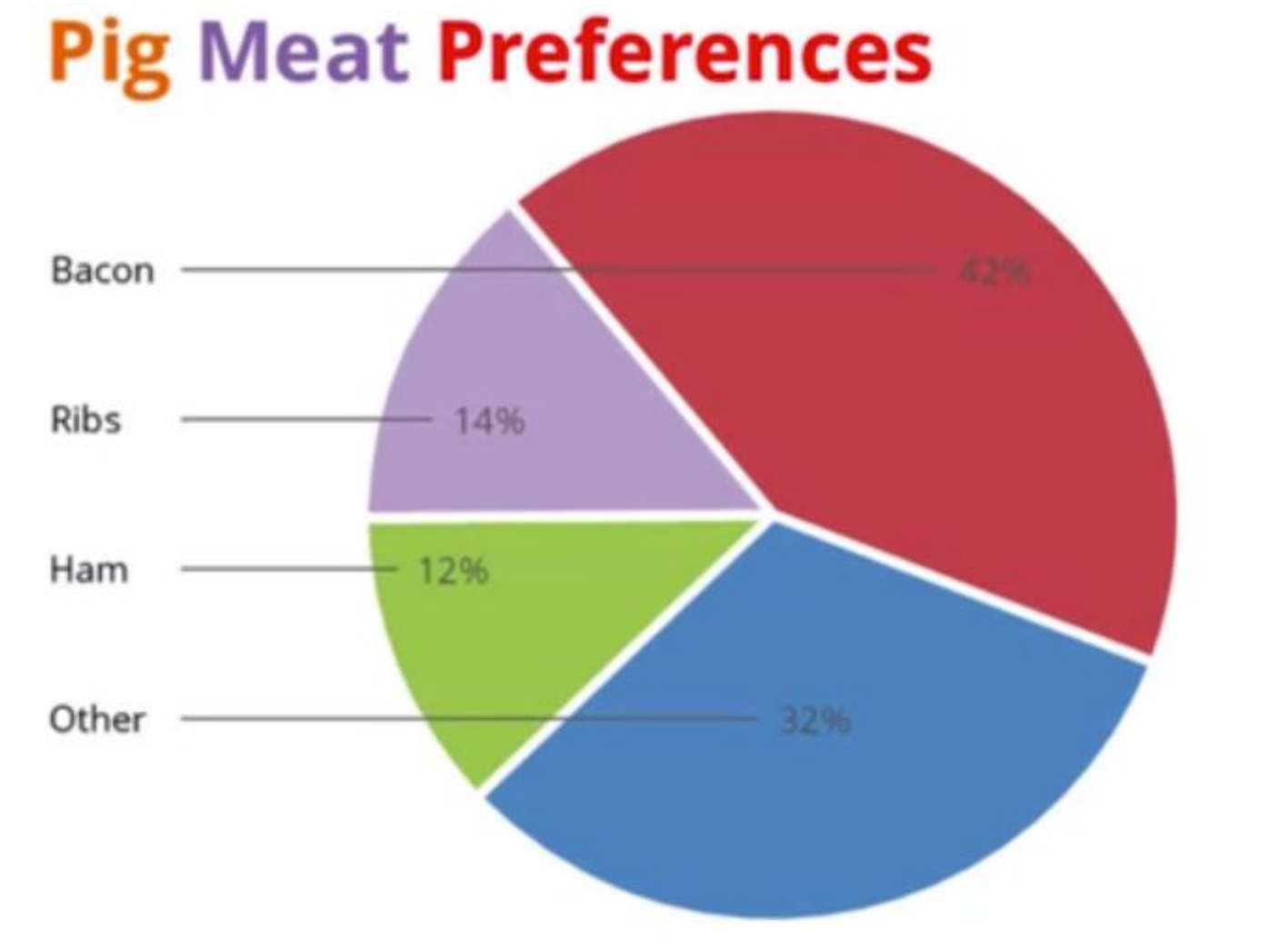
Visualization Example (6/12)



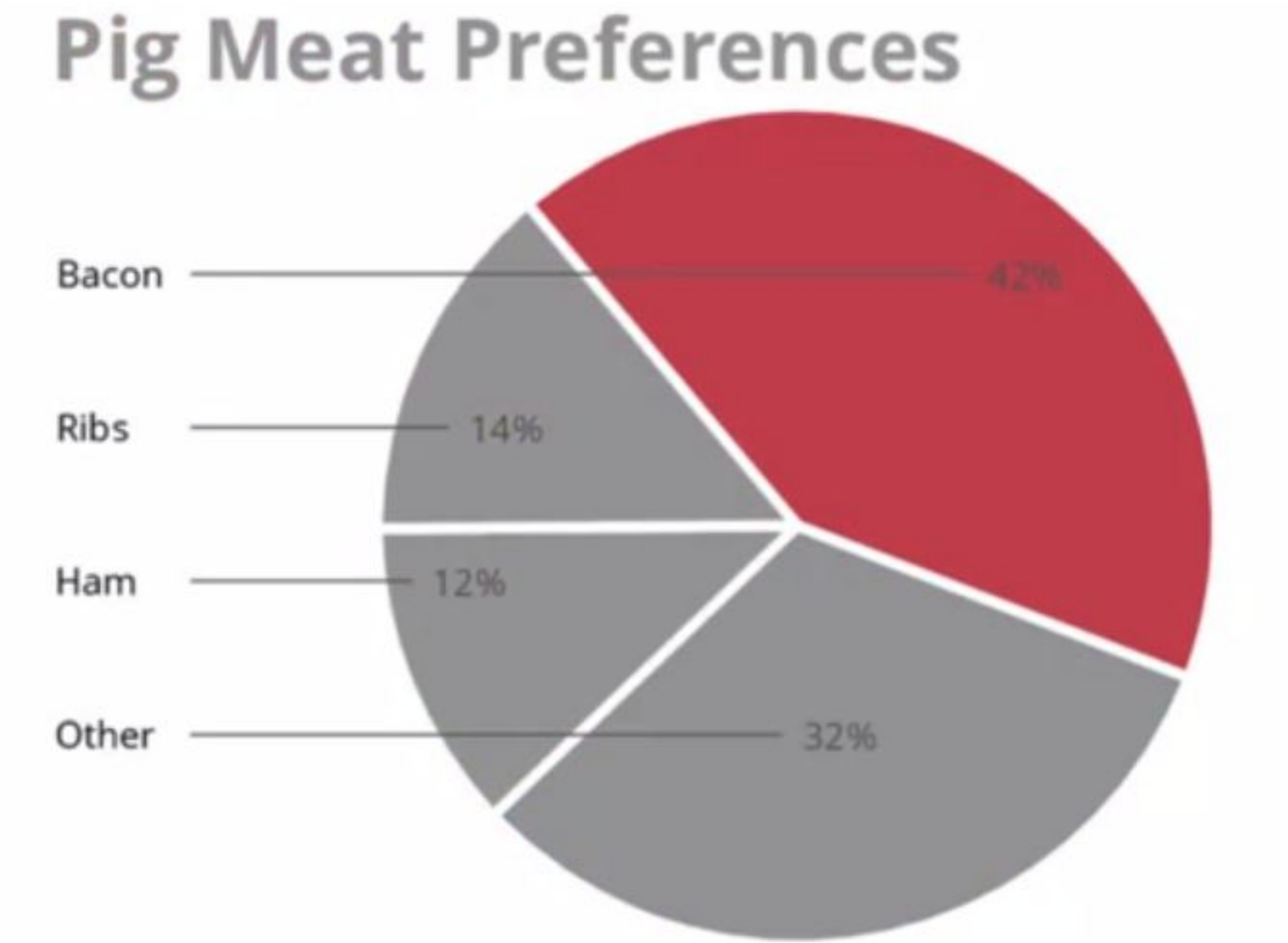
Visualization Example (7/12)



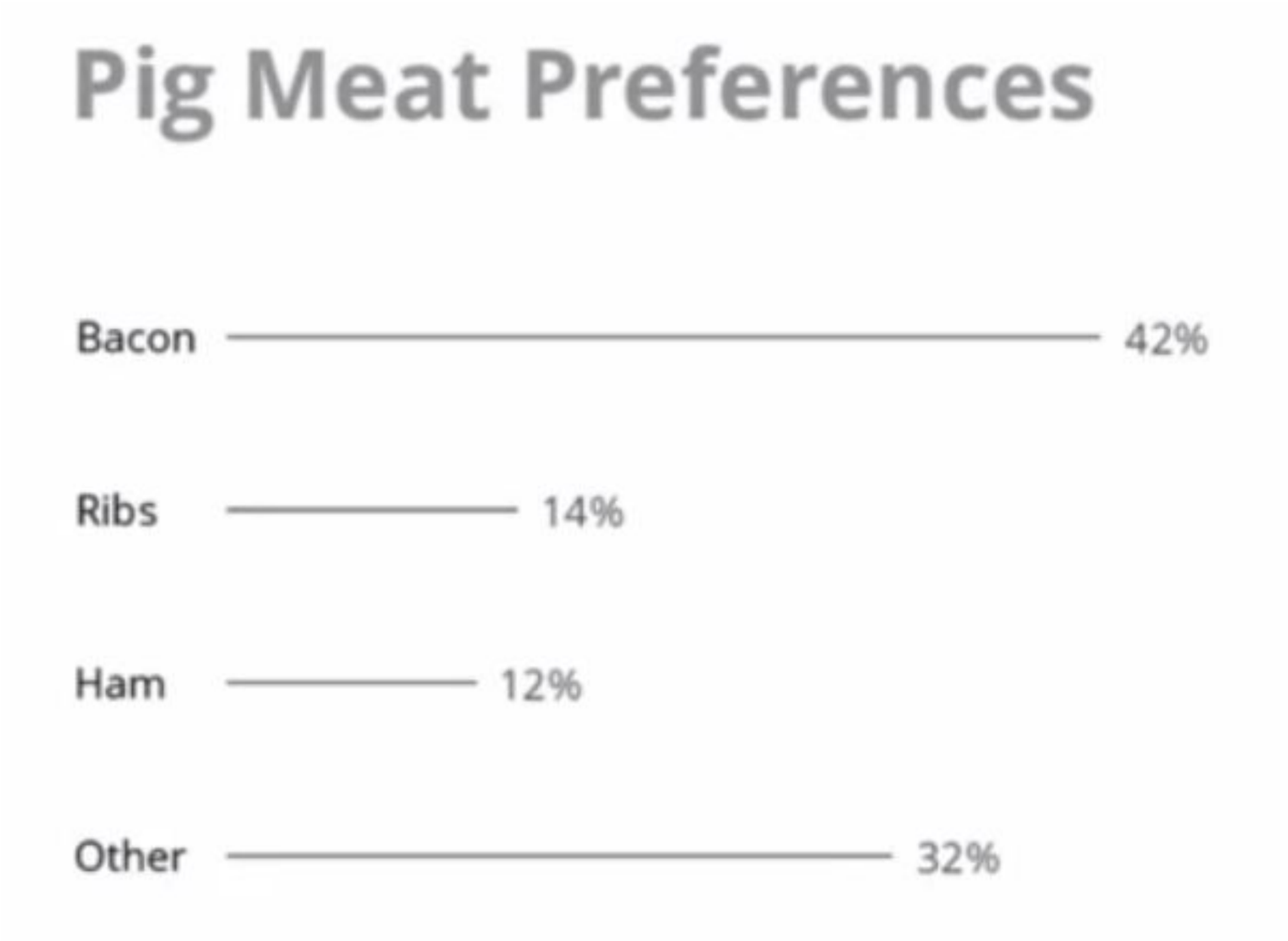
Visualization Example (8/12)



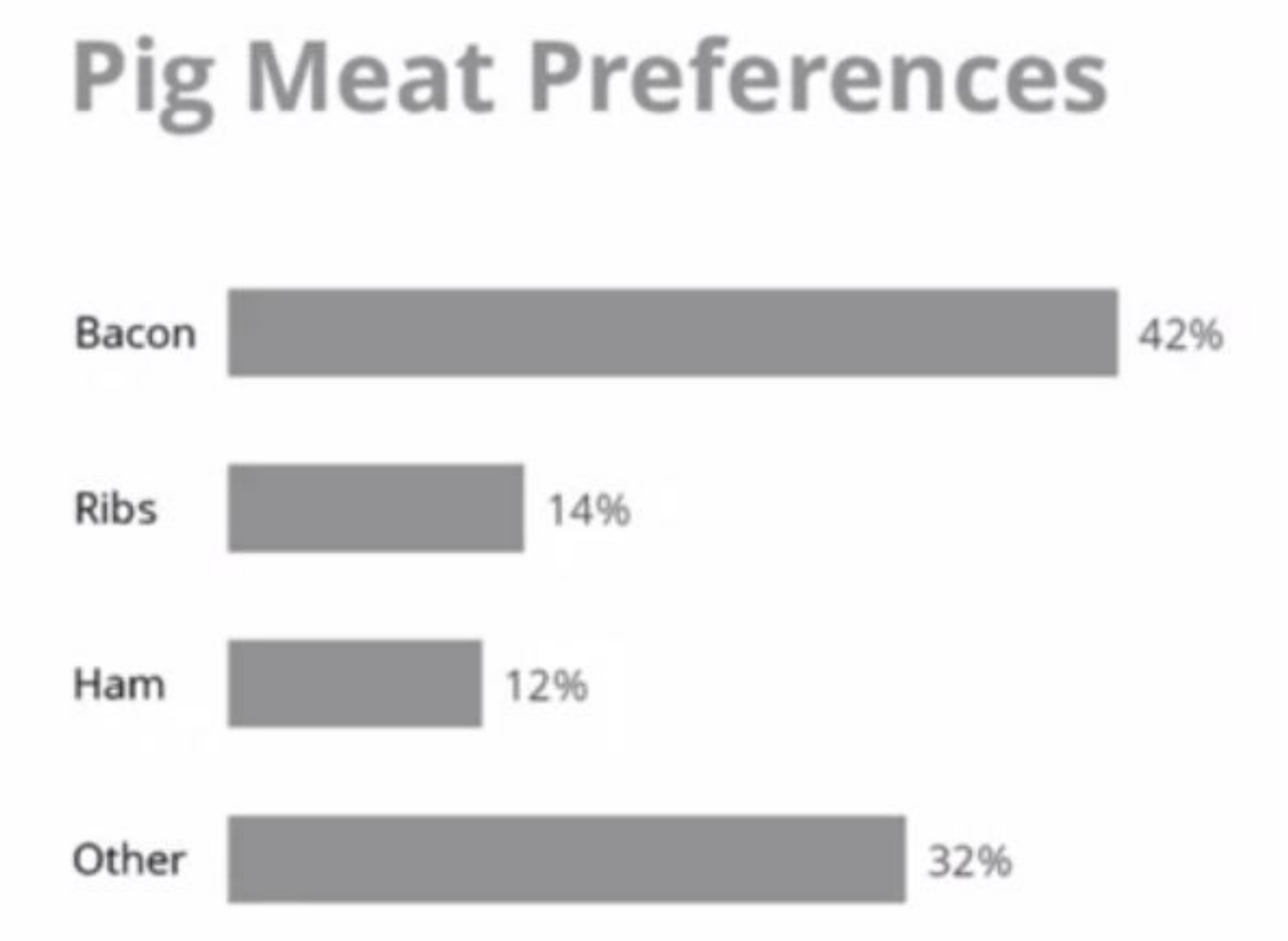
Visualization Example (9/12)



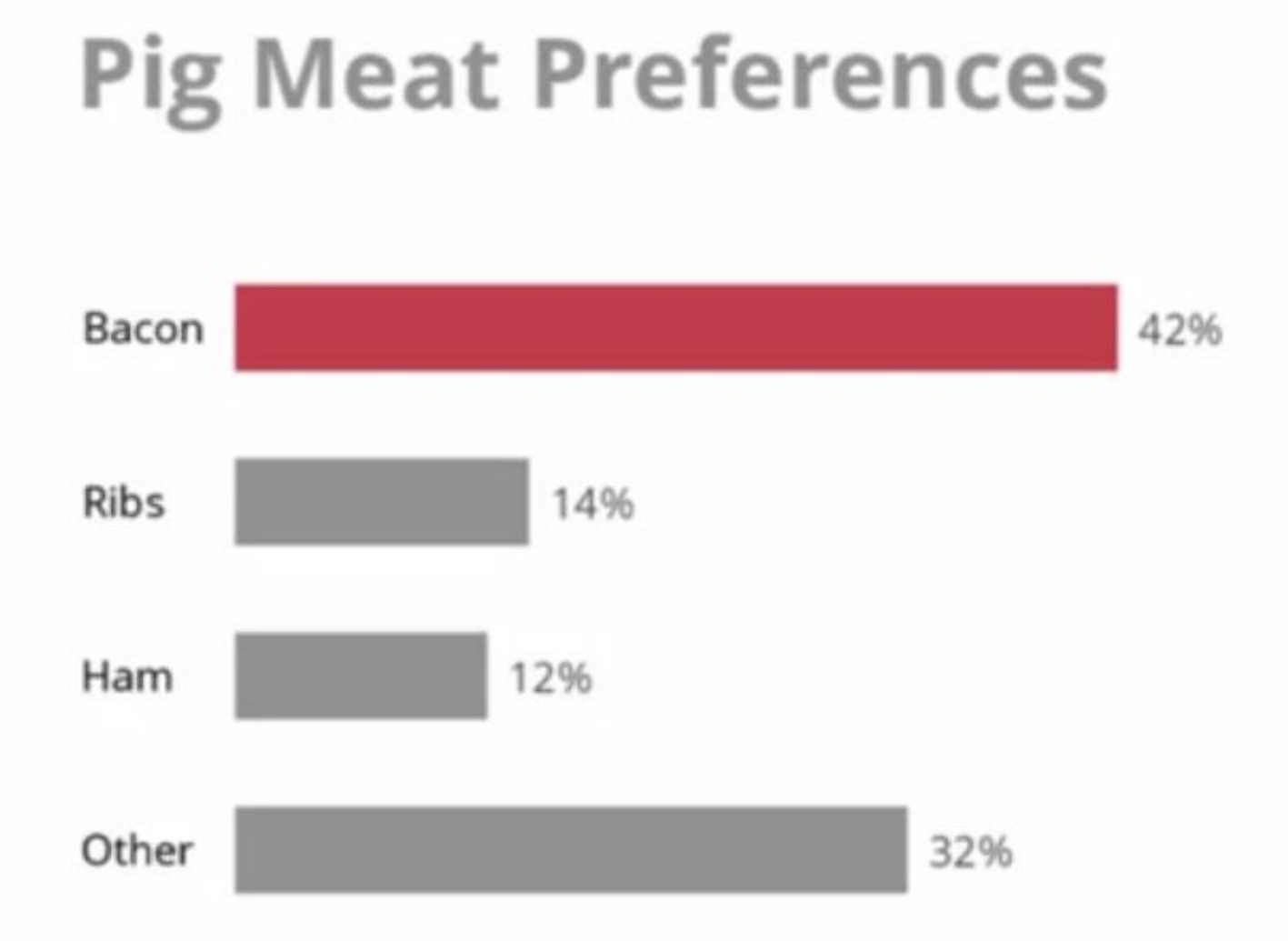
Visualization Example (10/12)



Visualization Example (11/12)

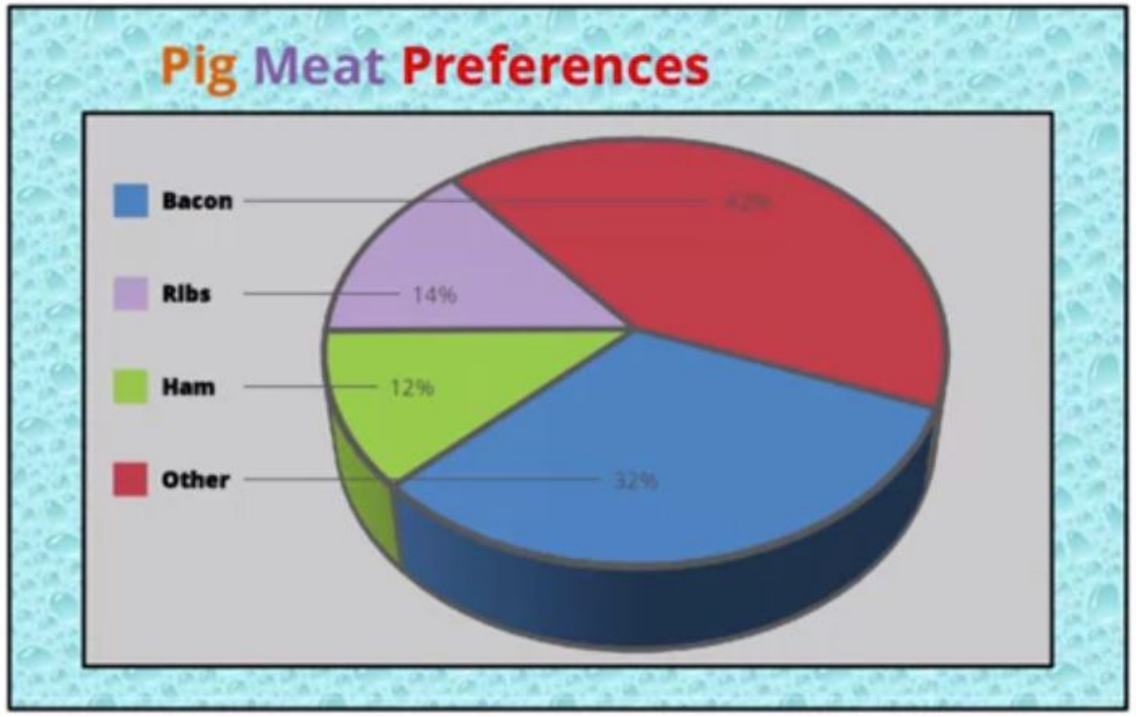


Visualization Example (12/12)

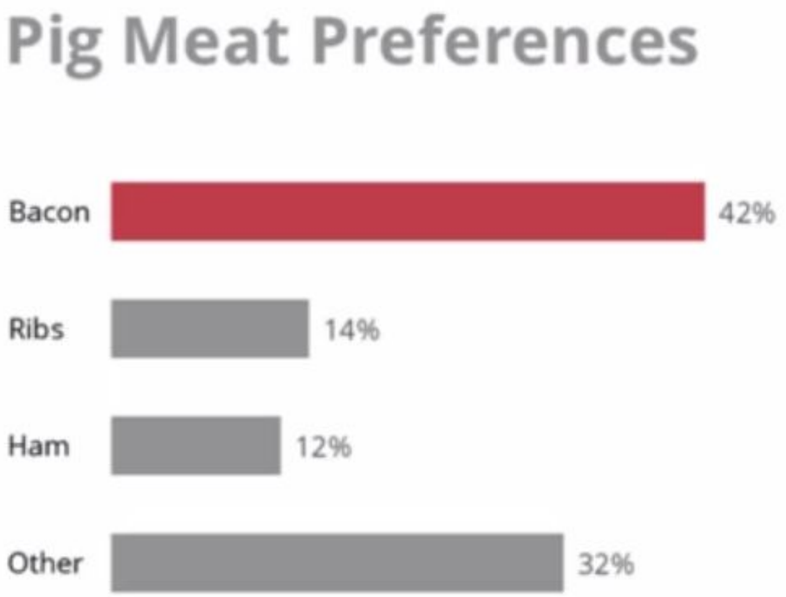


Visualization Example

Before



After



Google Sheets Charts

Questions

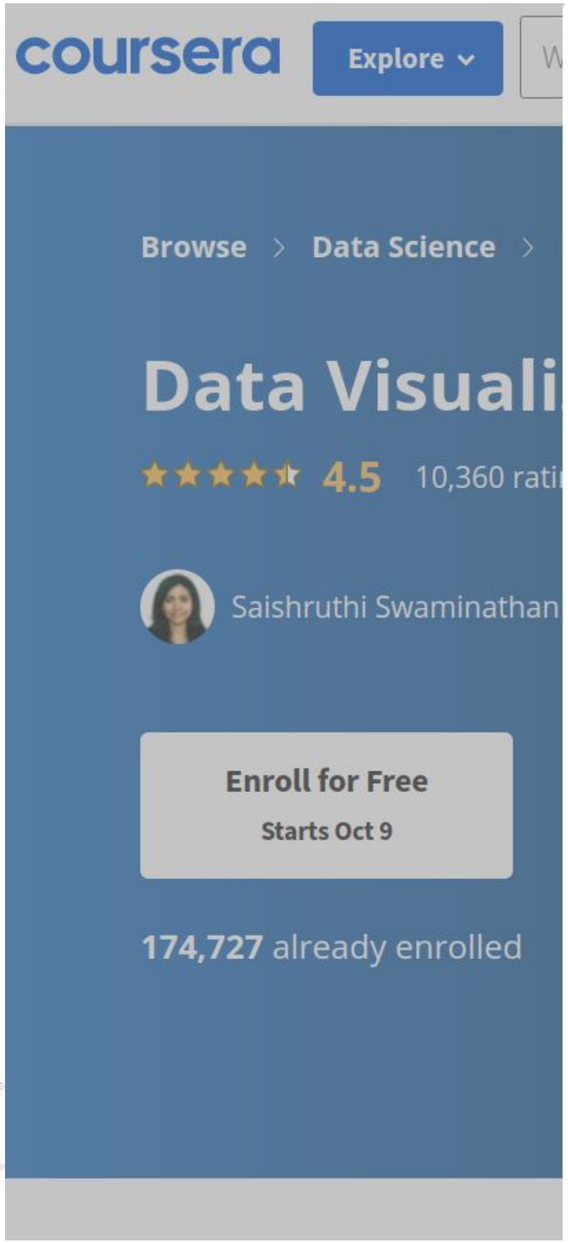
Links

https://github.com/fcai-b/dv

References

1. https://www.coursera.org/learn/foundations-data
2. https://www.coursera.org/learn/what-is-datascience
3. https://www.coursera.org/learn/python-for-data-visualization
4. https://www.coursera.org/learn/google-sheets---advanced-topics

Enroll in the Coursera courses



Step 2 of 2

7-day Free Trial

Data Visualization with Python is part of the larger IBM Data Science Professional Certificate.
Your 7-day free trial includes:

- ✓ **Unlimited access to all courses in the Certificate**
  Watch lectures, try assignments, participate in discussion forums, and more.
- ✓ **Cancel anytime.**
  No penalties - simply cancel before the trial ends if it's not right for you.
- ✓ **$39 USD per month to continue learning after trial ends.**
  Go as fast as you can - the faster you go, the more you save.
- ✓ **Certificate when you complete.**
  Share on your resume, LinkedIn, and CV.

Start Free Trial

Audit the course