



1



2



## Introduction to Statistics

Shep Sheppard  
Customer Engineer  
Fast Track ISV



The slide features a background with a network of interconnected nodes and lines in shades of orange, purple, and blue. The Microsoft logo is in the top left corner. The title 'Introduction to Statistics' is prominently displayed. Below the title, the speaker's name and role are listed. A portrait of Shep Sheppard is on the right side of the slide.

3

## Session goals

- Intro to Common Python Packages used for Data Engineering and Machine Learning
  - Pandas
  - Numpy
  - Seaborn/Matplotlib/Pillow
  - SciKit-Learn

4

4

## Pandas

- Name is derived from "Panel Data"
- Pandas is a python Package that aims to make data engineering fast, flexible and intuitive
- Support for multiple data types, read, write
- Perform statistical analysis
- Basic visualizations

- Notebook = 00-Pandas.ipynb



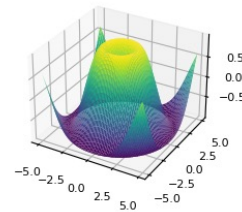
5

5

## Numpy

- Numerical Python
- Fundamental package for scientific computing in Python
- Provides a multidimensional array object support

- Notebook = 00-Numpy.ipynb



6

6

## SciKit-Learn

- scikit-learn is a Python module for machine learning and data engineering built on top of SciPy
- Classification
- Regression
- Clustering
- Dimensionality Reduction
- Data Preprocessing



7

7

## Seaborn/Matplotlib/Pillow

- **Seaborn** is a Python visualization library based on matplotlib
- Matplotlib is a comprehensive library for creating static, animated and interactive visualizations in python.
- Pillow is "Python Imaging Library"



[Notebook - 01-Visualization-The-Gateway-Drug.ipynb](#)  
[Notebook - 03-Seaborn.ipynb](#)

8

8

## Key learnings and insights

- Learnings and insights
  - Learnings and insights 1
  - Learnings and insights 2

This slide is required—**do not delete**—please read the notes for this slide, then **delete this text box**

9

9

## Thank you for attending the MLADS Conference and helping to build a strong community

To find recordings, presentations, and other resources from the event, go to: <http://aka.ms/november2022mlads>.

More information about the AI & ML Connected Community: <http://aka.ms/aiml-cc>.

### We want your feedback!

We read each evaluation and incorporate your comments into future MLADS events. Both sessions and event evaluations will be available within the MyMLADS event portal during the conference dates—to access them, navigate to the [Evaluations tab](#).

We request that session evaluations are submitted following each day of programming.

10

10


A presentation slide with a dark blue background featuring a light blue network pattern of interconnected dots and lines. The text "Q&A" is displayed in white on the left side. A white rectangular box with a dashed border is centered on the slide, containing the text: "This slide is required—do not delete—please read the notes for this slide, then delete this text box". The number "11" is located in the bottom left corner of the slide.

Q&A

This slide is required—do not delete—please read the notes for this slide, then delete this text box

11

11

A presentation slide with a solid dark blue background. The Microsoft logo, consisting of four colored squares (red, green, blue, yellow) followed by the word "Microsoft", is positioned in the top left corner. At the bottom left, there is a small line of text: "© Copyright Microsoft Corporation. All rights reserved."

Microsoft

© Copyright Microsoft Corporation. All rights reserved.

12