

Current and future tools

One of the greatest skills a data professional can have is learning how to apply their knowledge of one tool to another tool. Throughout your career, you might discover that different organizations you work for use different tools—and in the field of data science, new and emerging technologies mean that exciting new tools are being developed all the time. This means there will always be opportunities to expand your data science toolkit! In this reading, you will learn more about tools today, including some of the tools you're going to learn about in this program. You will also explore some of the exciting ways tools are evolving and what that might mean for your toolkit in the future. Finally, you will explore a demonstration that illustrates how you'll be using some of these tools in the very near future.

Tools today

In this certificate program, you will have the opportunity to learn about many common tools data professionals use every day: spreadsheets, databases, query languages, data visualization, programming languages, and dashboards. Understanding the current tool landscape—and how it's changing—will help you continue to grow your data science skills throughout your career. And understanding how the skills you learn for one tool can be applied to another means that you can adapt and add more tools to your toolkit!

Tool	Definition	Examples	Transferable skills
Spreadsheets	A digital worksheet where data can be manipulated and used for calculations	<ul style="list-style-type: none">• Google Sheets• Microsoft Excel	<ul style="list-style-type: none">• Data entry• Mathematical calculations• Manage datasets• Task automation• Data manipulation• Data analysis
Databases	A collection of data stored in a computer system	<ul style="list-style-type: none">• Google Cloud• CloudSQL• Oracle• Microsoft SQL Server	<ul style="list-style-type: none">• Database design• Data storage management• Data integrity
Programming languages	A system of words and symbols used to write instructions that computers follow	<ul style="list-style-type: none">• SQL• R• Python• Java• C++	<ul style="list-style-type: none">• Communicate with computer systems• Write and input commands• Manage datasets• Data manipulation• Data analysis
Data visualization	The graphical representation of data	<ul style="list-style-type: none">• Tableau• Matplotlib• Seaborn• Google Charts• InfoGram	<ul style="list-style-type: none">• Communicate data insights• Design compelling visuals

Tool	Definition	Examples	Transferable skills
Dashboards	A tool that monitors live, incoming data	<ul style="list-style-type: none"> • ChartBlocks 	<ul style="list-style-type: none"> • Identify key metrics
		<ul style="list-style-type: none"> • Tableau • LookerStudio • Microsoft PowerBI 	<ul style="list-style-type: none"> • Communicate data insights • Monitor real-time data • Develop data visualizations • Design filters and custom calculations

Already, there are so many tools to choose from as a data professional. This certificate program will focus primarily on Python and data visualizations. As you progress in your career, you might find yourself learning new tools, and using your existing skills in new ways. Being able to recognize where tool skills overlap will help you continuously grow your data toolkit now and in the future.

You in the near future

So far in this reading, you have been considering how the skills you're going to learn in this certificate program will help you use even more tools in the future. As you prepare for your learning journey, you can also think about how you'll be able to apply these skills soon—not just in the distant future.

This certificate program focuses on some of the most commonly used tools for data analytics and machine learning with Python. More specifically, you will use:

- NumPy and pandas for data processing and manipulation
- matplotlib.pyplot, seaborn, and Tableau for visualizations
- statsmodels for statistical tests and regression modeling
- scikit-learn for building machine learning models

Next, consider the following overview of some of the tools you'll use to complete tasks in this certificate program.

You'll use pandas to view and manipulate tabular data with Python. In the following example, a comma-separated value (.csv) file is read into a pandas dataframe, of which the first five rows are displayed. A dataframe is basically a table used to organize data. This data is from the [UC Irvine Machine Learning Repository](#). It contains the count of public bicycles rented per hour in the Seoul Bike Sharing System, with corresponding weather data and holiday information.

```
# Read in the data
df = pd.read_csv('SeoulBikeData.csv', encoding='latin1')
df.head()
```

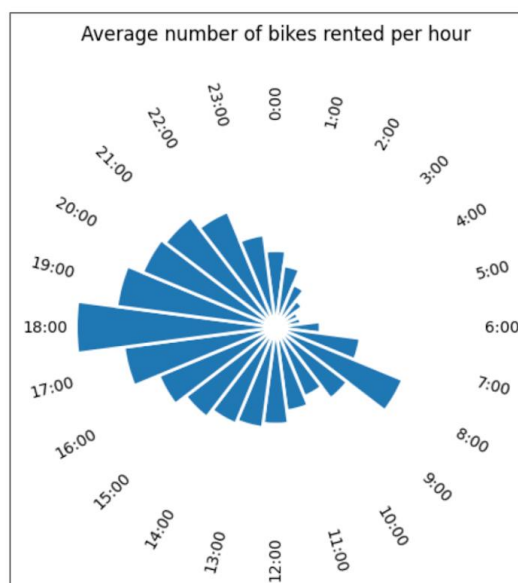
	Date	Rented Bike Count	Hour	Temperature(°C)	Humidity(%)	Wind speed (m/s)	Visibility (10m)	Dew point temperature(°C)
0	01/12/2017	254	0	-5.2	37	2.2	2000	-17.6
1	01/12/2017	204	1	-5.5	38	0.8	2000	-17.6
2	01/12/2017	173	2	-6.0	39	1.0	2000	-17.7
3	01/12/2017	107	3	-6.2	40	0.9	2000	-17.6
4	01/12/2017	78	4	-6.0	36	2.3	2000	-18.6

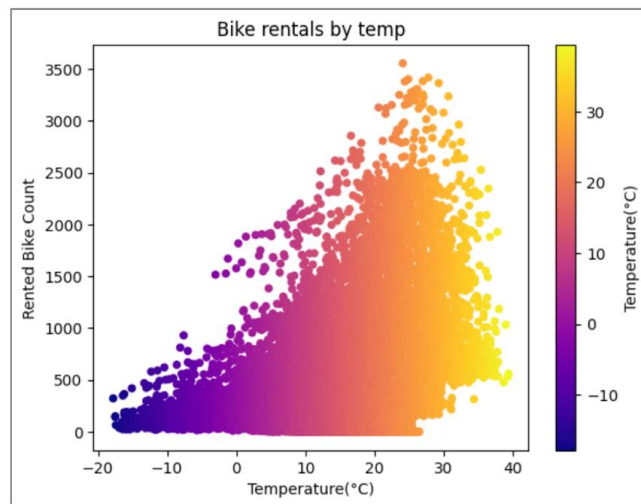
You'll use NumPy and pandas to perform calculations and get statistics for your data.

```
# Check the statistics
df.describe()
```

	Rented Bike Count	Hour	Temperature(°C)	Humidity(%)	Wind speed (m/s)	Visibility (10m)	Dew point temperature(°C)
count	8760.000000	8760.000000	8760.000000	8760.000000	8760.000000	8760.000000	8760.000000
mean	704.602055	11.500000	12.882922	58.226256	1.724909	1436.825799	4.073813
std	644.997468	6.922582	11.944825	20.362413	1.036300	608.298712	13.060369
min	0.000000	0.000000	-17.800000	0.000000	0.000000	27.000000	-30.600000
25%	191.000000	5.750000	3.500000	42.000000	0.900000	940.000000	-4.700000
50%	504.500000	11.500000	13.700000	57.000000	1.500000	1698.000000	5.100000
75%	1065.250000	17.250000	22.500000	74.000000	2.300000	2000.000000	14.800000
max	3556.000000	23.000000	39.400000	98.000000	7.400000	2000.000000	27.200000

You'll use Tableau, matplotlib.pyplot, and seaborn to create data visualizations.

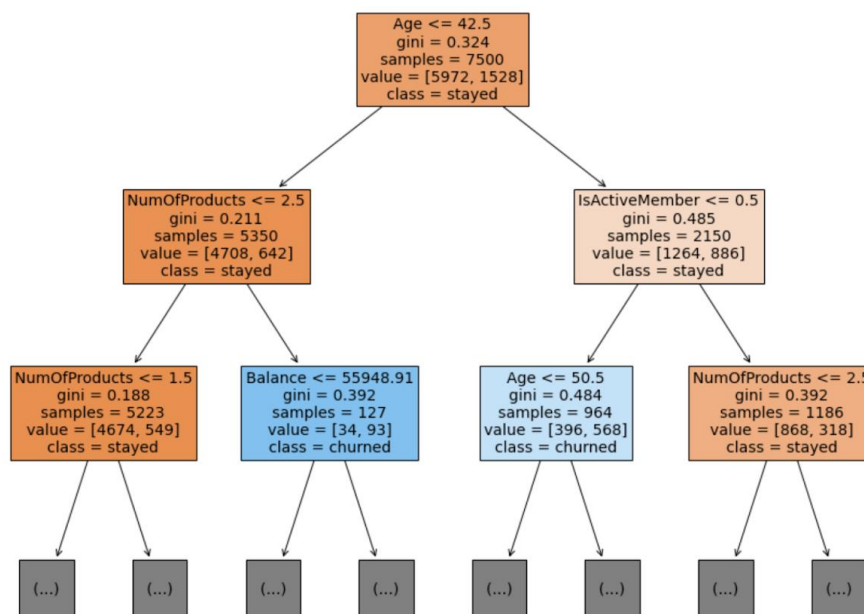




You'll use statsmodels to practice statistical analysis and linear regression.

OLS Regression Results						
Dep. Variable:	Sales	R-squared:	0.871			
Model:	OLS	Adj. R-squared:	0.870			
Method:	Least Squares	F-statistic:	1918.			
Date:	Tue, 09 Aug 2022	Prob (F-statistic):	1.38e-253			
Time:	20:43:11	Log-Likelihood:	-2798.9			
No. Observations:	572	AIC:	5604.			
Df Residuals:	569	BIC:	5617.			
Df Model:	2					
Covariance Type: nonrobust						
	coef	std err	t	P> t	[0.025	0.975]
Intercept	300.8532	2.512	119.789	0.000	295.920	305.786
C(TV)[T.Low]	-209.8691	3.394	-61.841	0.000	-216.535	-203.203
C(TV)[T.Medium]	-105.4952	3.379	-31.224	0.000	-112.131	-98.859
Omnibus:	547.584	Durbin-Watson:	1.960			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	37.103			
Skew:	0.015	Prob(JB):	8.77e-09			
Kurtosis:	1.753	Cond. No.	3.97			

And you'll use scikit-learn to build and analyze machine learning models:



This is just a small sample of the full range of topics you'll learn about in this certificate. As you gain proficiency with these tools, you'll be equipped to take on nearly any data task.

Tools tomorrow and beyond

The world of data science is always growing and evolving—tools you might not have even known about a few years ago can quickly become necessary for professionals working in the field. As you consider the skills you are developing now, it can be useful to consider all the ways you might also use them in the future.

Artificial intelligence

Artificial intelligence, or AI, refers to computer systems that are able to perform tasks that normally require human intelligence. One of the great benefits of using AI for data science is that it can help provide real-time insights to stakeholders. For example, a business with an e-commerce website might use AI to monitor and deliver insights about how customers use their website in real-time, allowing the team to make quick improvements.

Machine learning

Machine learning is the use and development of algorithms and statistical models to teach computer systems to analyze and discover patterns in data. Data analysts can train algorithms to analyze large amounts of data that would otherwise take a long time to process. For example, a financial analyst might use machine learning to find patterns in the data that help identify fraud.

One of the most exciting developments in these future technologies is the way they can be used together to automate tasks and provide insights faster than ever.

Key takeaways

As a data professional, you will continue learning new skills and applying your current skills in new ways throughout your career. Recognizing how skills can be transferable allows you to adapt to different organizations' needs and evolving technologies. And as you progress through this, you add tools to your data science toolbox that will help you now and in the future!