March 2024

STATISTIC PROBABILIY IN PYTHON

Presented by TIMOTHEE NKWAR

INSTRUCTOR : Dr. Aysegul Erem

1. PYTHON

Python is a programming language that’s:

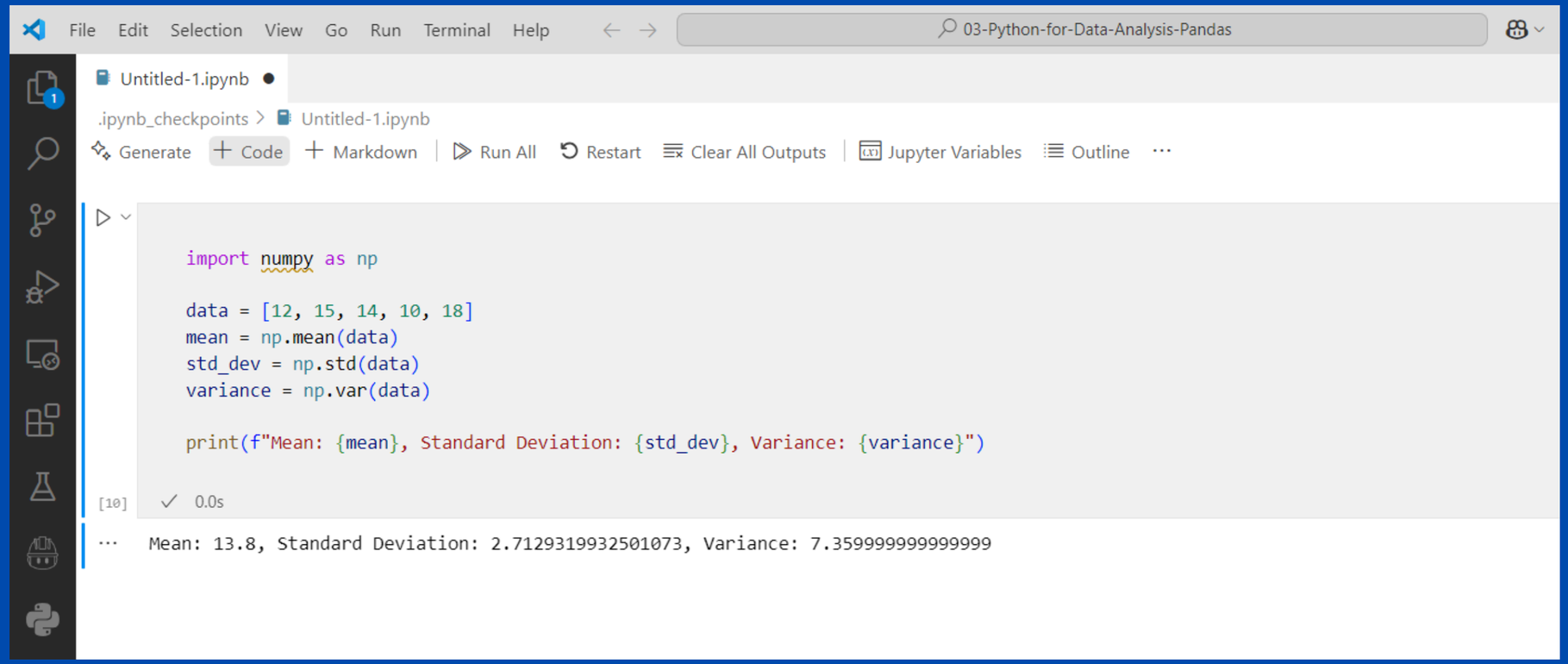
✅ Easy to read and learn.
✅ Versatile — used for Data Science, Web Development, Automation, AI, Games, and more.
✅ Backed by a massive community with thousands of ready-to-use libraries.

SOME PYTHON LIBRARIES FOR STATISTICS?

Libraries	What it's for
Numpy	Mathematical & statistical operations, arrays, basic stats (mean, median, variance) .
scipy.stats	Advanced statistics, distributions, hypothesis testing.
Pandas	Data manipulation + summary statistics.
matplotlib & seaborn	Visualization of distributions, boxplots, histograms.

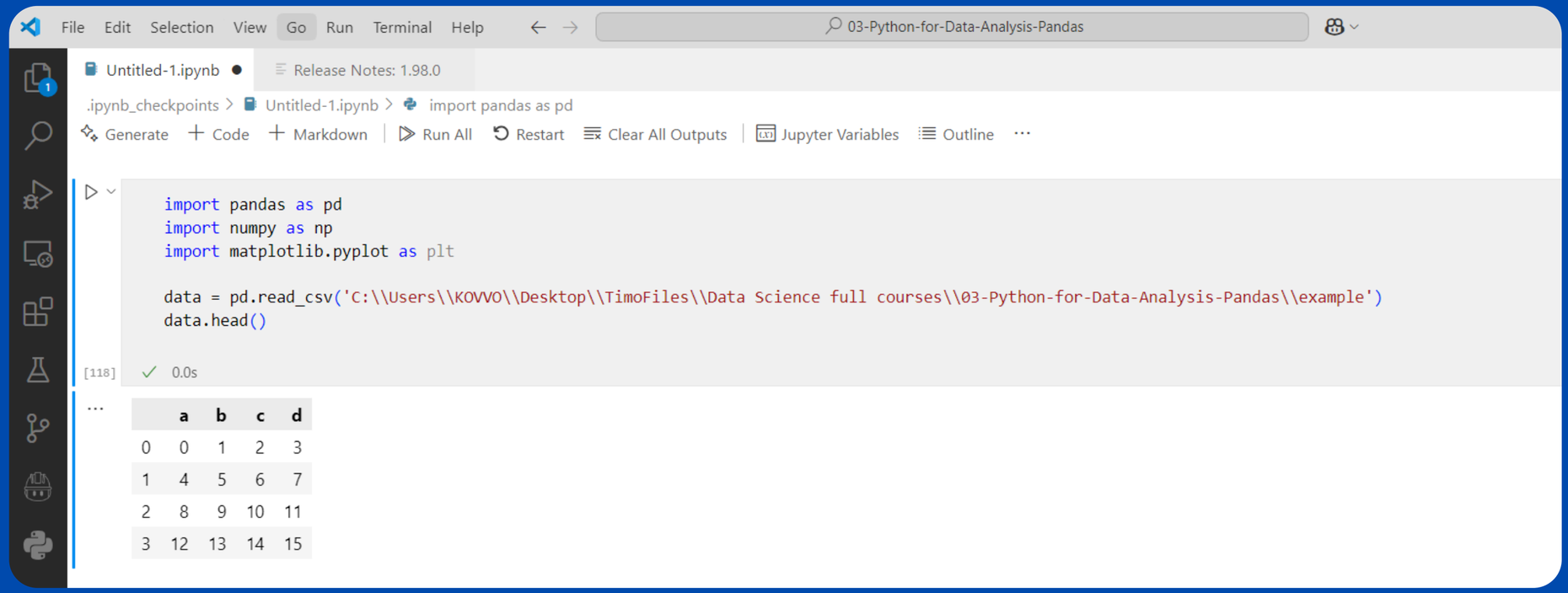
a. NumPy - Core Stats

NumPy is your foundation for numerical and statistical operations:



c. pandas - Data Analysis

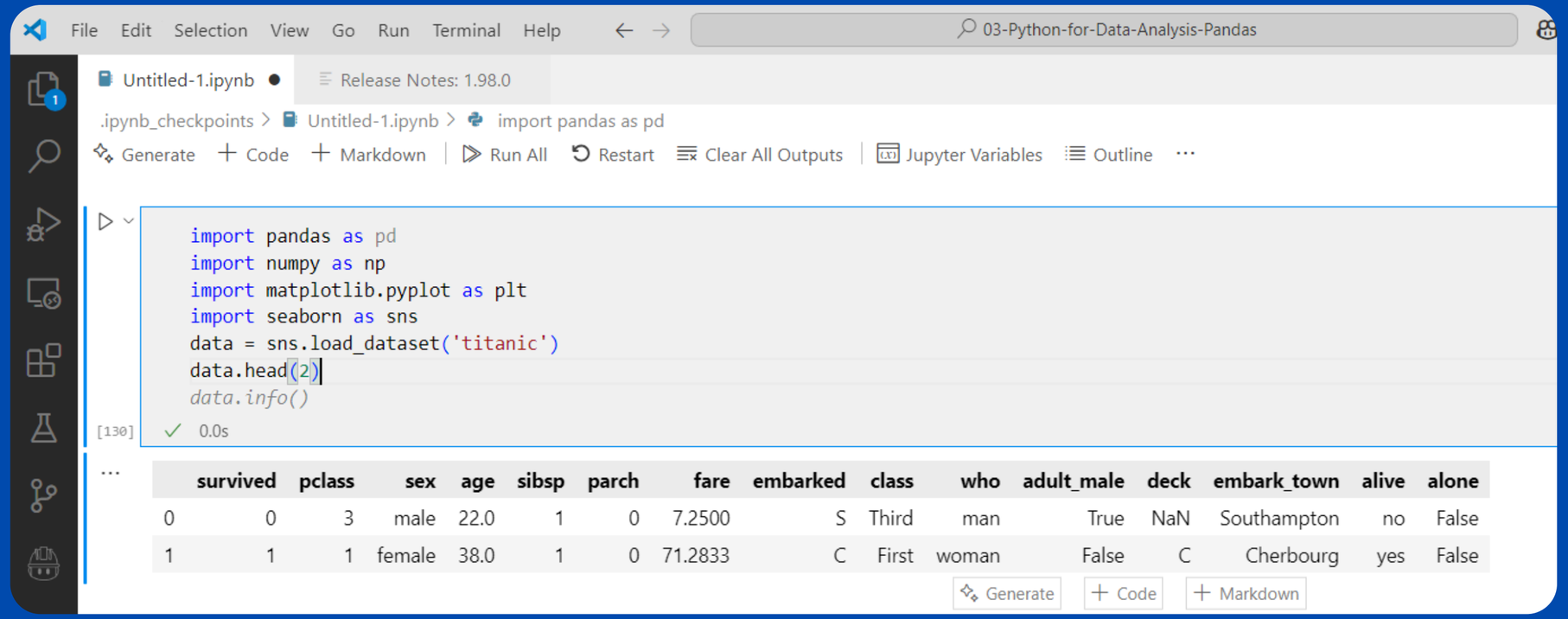
pandas is for working with dataframes and generating quick statistics:
This is how to load data with pandas .



	a	b	c	d
0	0	1	2	3
1	4	5	6	7
2	8	9	10	11
3	12	13	14	15

d. seaborn & matplotlib

For visualizing distributions, correlations , and trends , seaborn datasets :



	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no	False
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	False

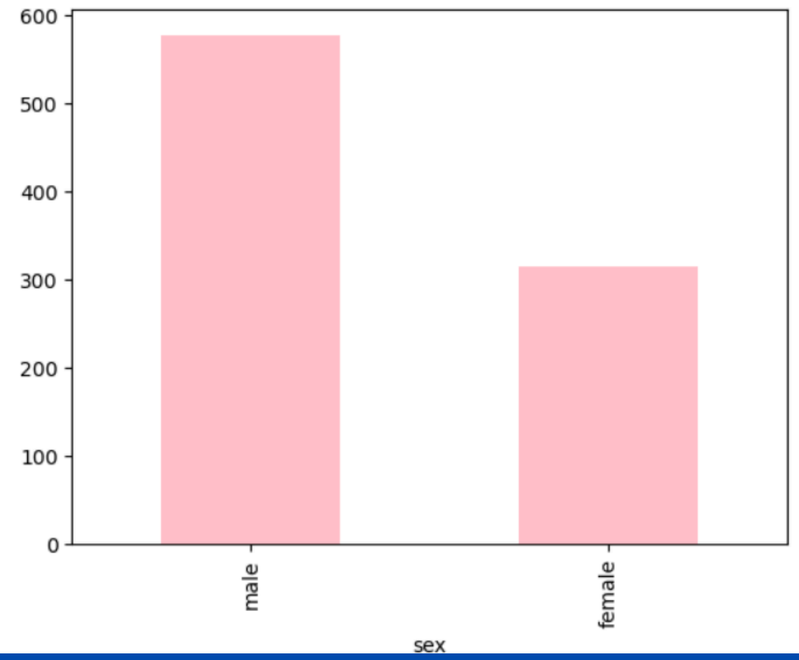
d.matplotlib - Visual Stats

```
gender = data["sex"].value_counts()

gender.plot(kind="bar",color = "pink" )
plt.figure(figsize=(8,6))
```

[132] ✓ 0.0s

```
<Figure size 800x600 with 0 Axes>
```



2. RANDOM VARIABLE

A RANDOM VARIABLE IS A VARIABLE WHOSE VALUES ARE DETERMINED BY THE OUTCOMES OF A RANDOM PROCESS OR EXPERIMENT. IT CAN TAKE DIFFERENT VALUES BASED ON CHANCE AND IS TYPICALLY ASSOCIATED WITH A PROBABILITY DISTRIBUTION.

THERE ARE TWO TYPES OF RANDOM VARIABLES:

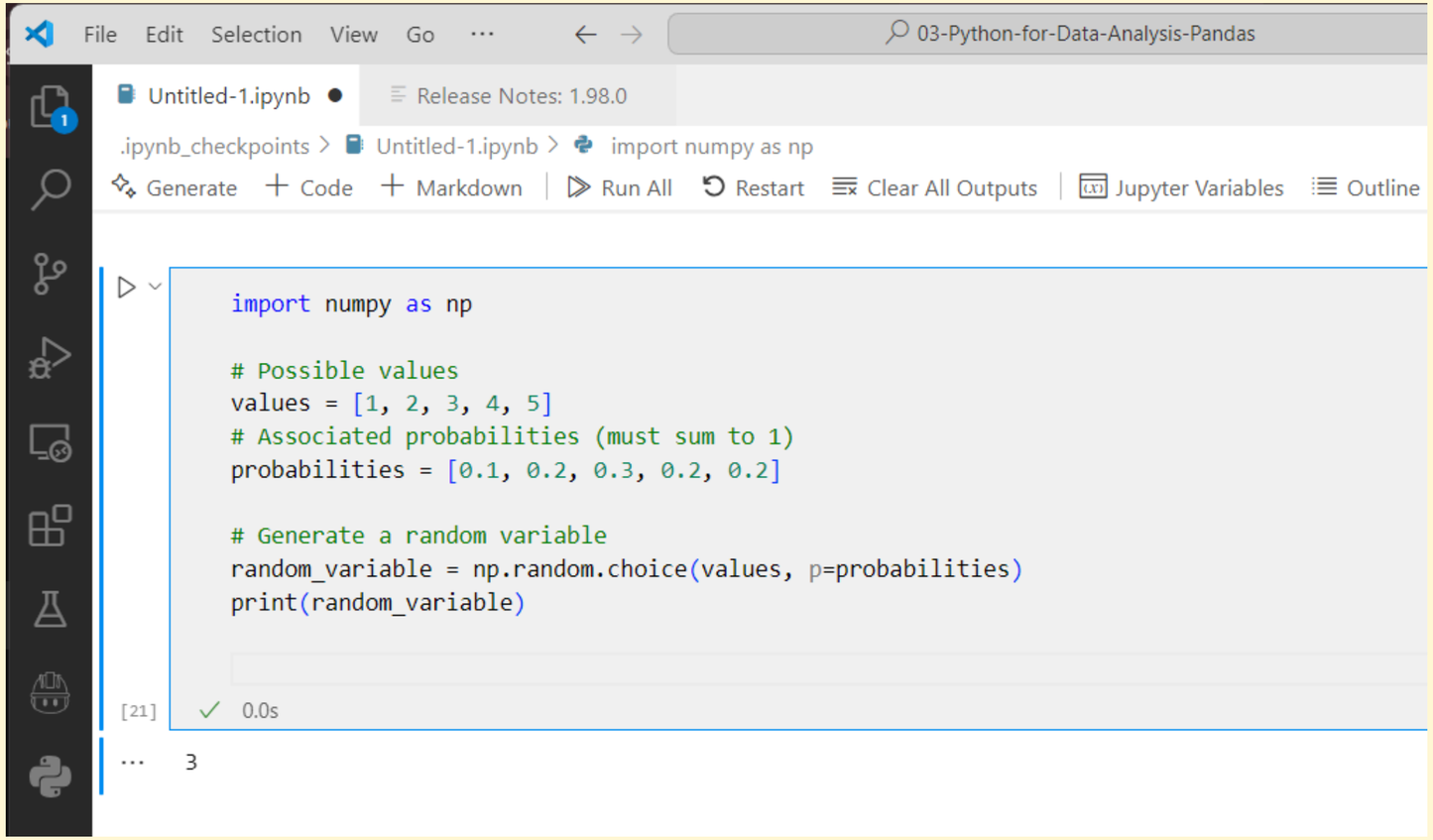
DISCRETE RANDOM VARIABLE: TAKES ON A FINITE OR COUNTABLE NUMBER OF VALUES. FOR EXAMPLE, THE NUMBER OF HEADS IN A COIN TOSS (EITHER 0 OR 1)

CONTINUOUS RANDOM VARIABLE: TAKES ON AN INFINITE NUMBER OF POSSIBLE VALUES WITHIN A GIVEN RANGE. FOR EXAMPLE, THE HEIGHT OF A PERSON, WHICH CAN BE ANY REAL NUMBER WITHIN A CERTAIN RANGE.

DISCRETE RANDOM VARIABLE

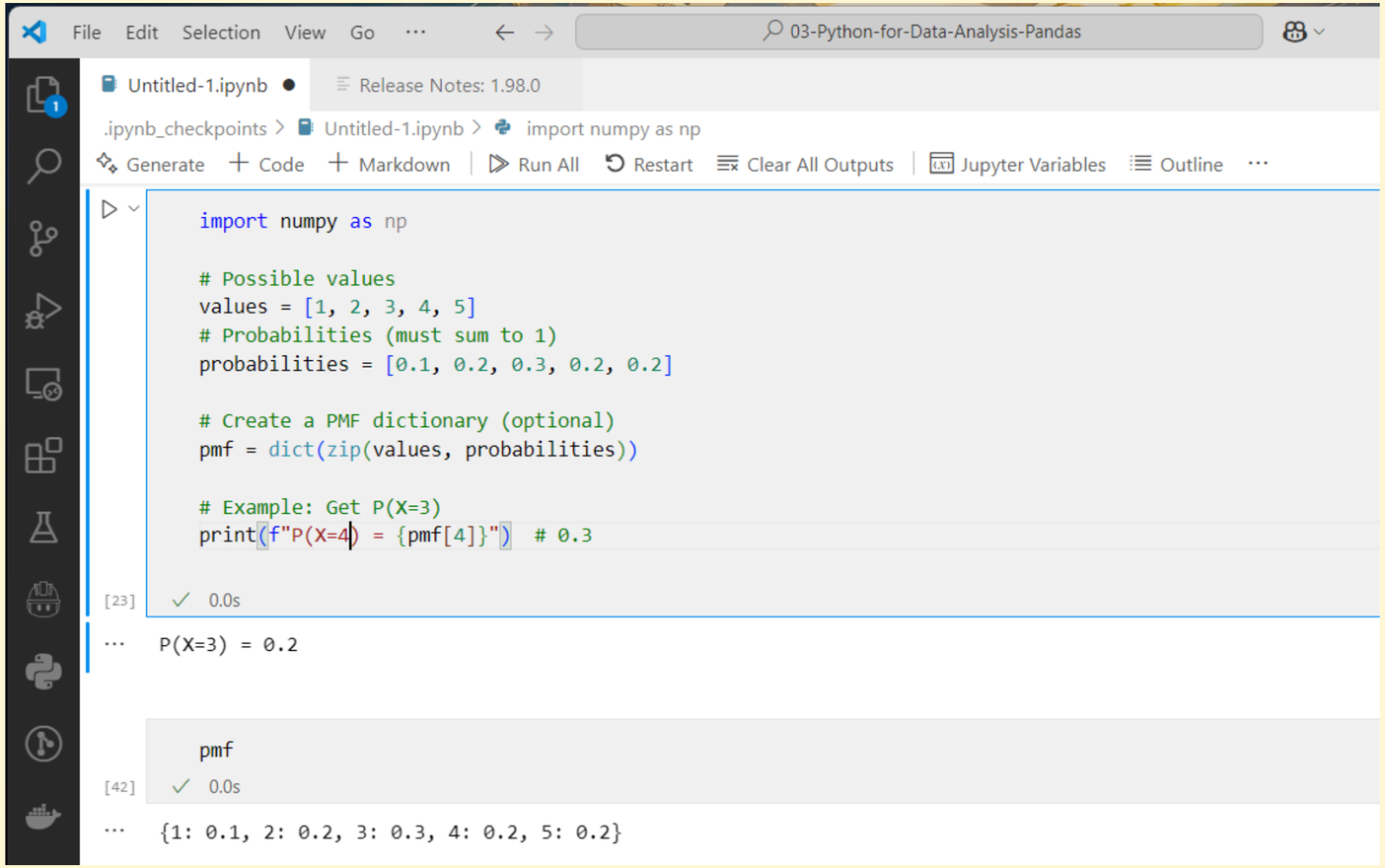
DISCRETE RANDOM VARIABLE: TAKES ON A FINITE OR COUNTABLE NUMBER OF VALUES. FOR EXAMPLE, THE NUMBER OF HEADS IN A COIN TOSS (EITHER 0 OR 1) .

DICE ROLL: {1, 2, 3, 4, 5, 6}



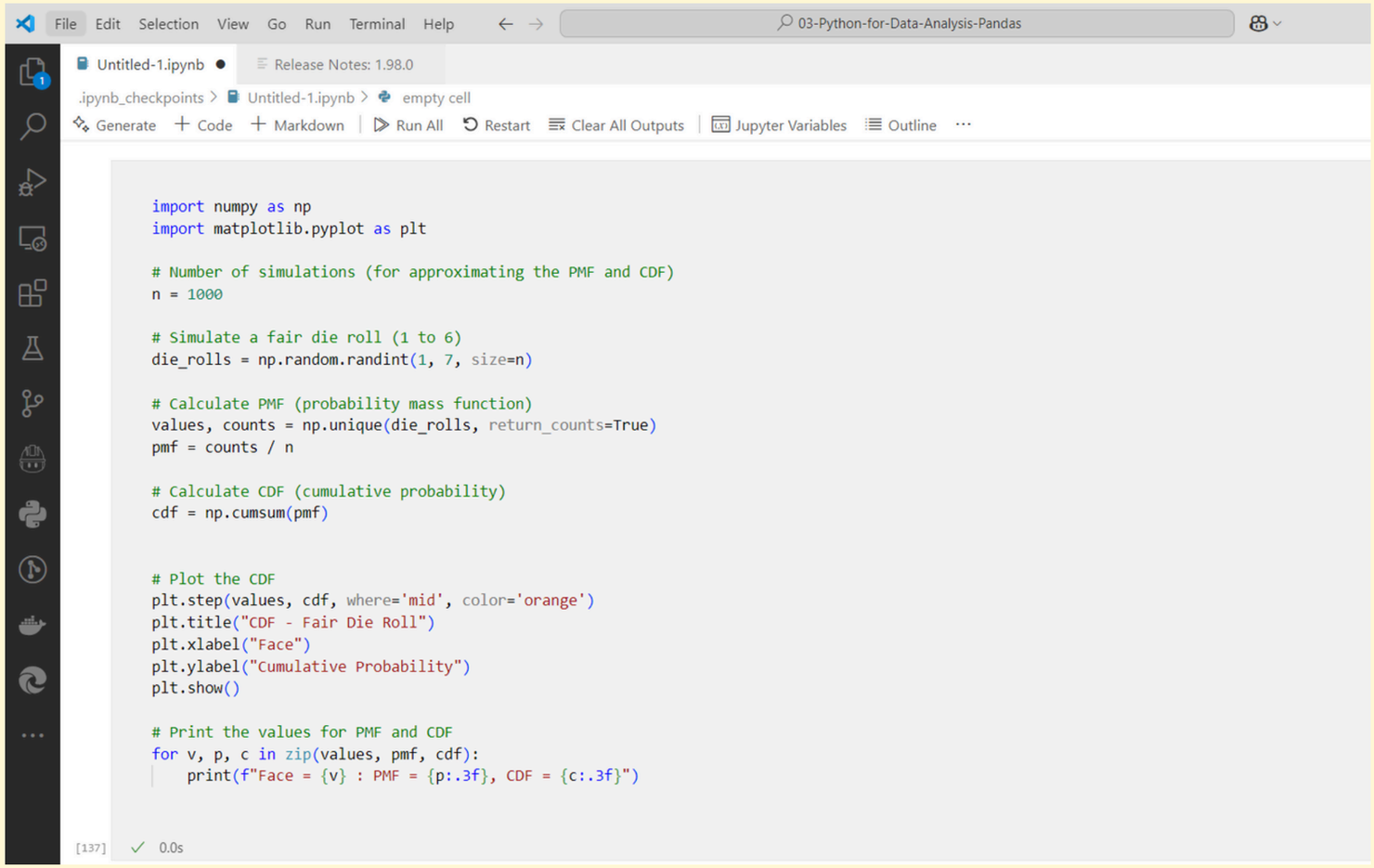
PMF OF DISCRETE RANDOM VARIABLE

if you want the probability of a specifique value like the P(X=3)
Here how to do that:

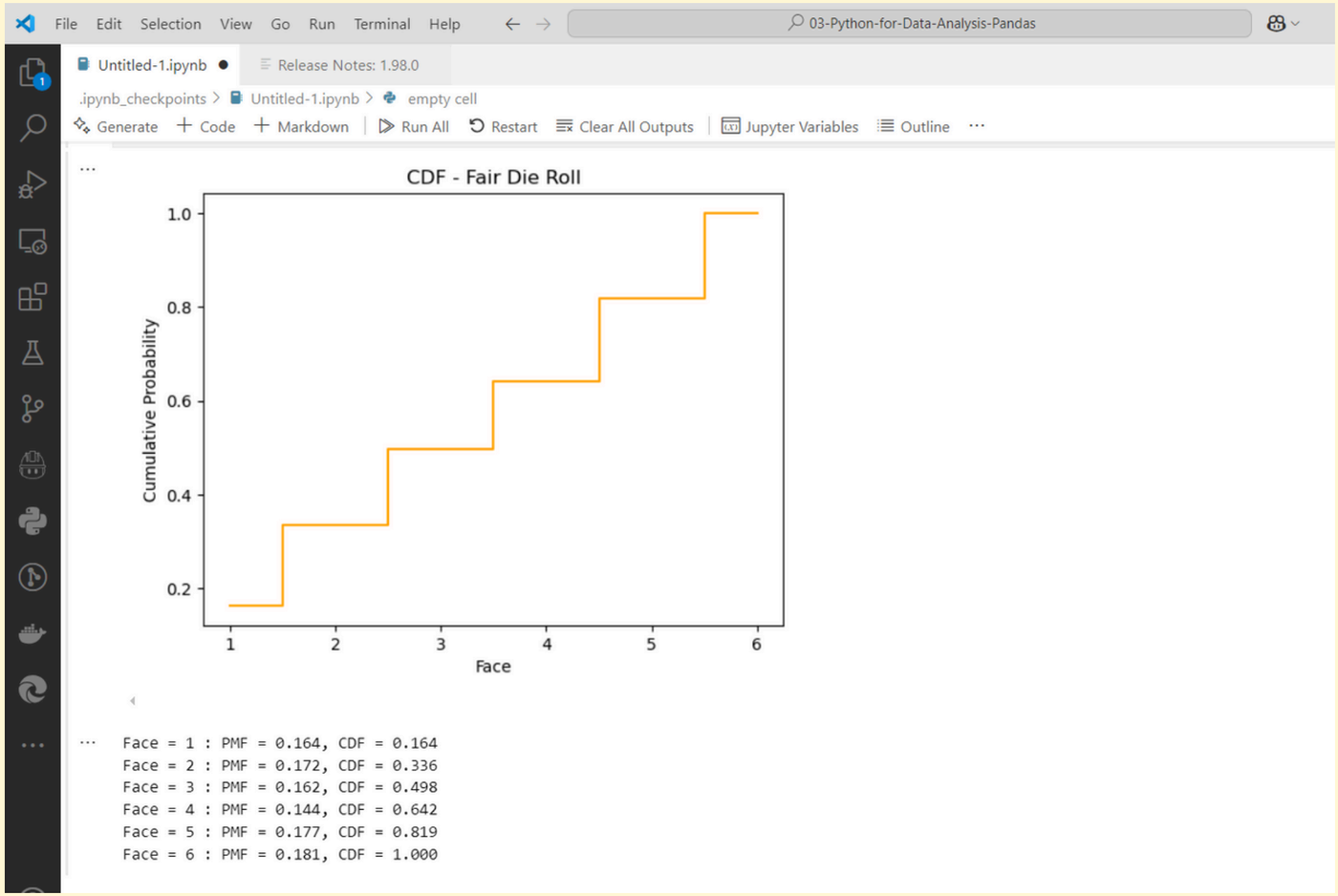


CDF OF DISCRETE RANDOM VARIABLE

It gives the cumulative probability of obtaining a value less than or equal to x.
For example, if you roll a fair 6-sided die, the CDF tells you the probability of rolling a number that is less than or equal to a given value.
If X=3, the CDF, F(3), would be the probability of rolling a 1, 2, or 3.



CDF OF DISCRETE RANDOM VARIABLE(CONT)



CONTINUOUS RANDOM VARIABLE

CONTINUOUS RANDOM VARIABLE: TAKES ON AN INFINITE NUMBER OF POSSIBLE VALUES WITHIN A GIVEN RANGE. FOR EXAMPLE, THE HEIGHT OF A PERSON, WHICH CAN BE ANY REAL NUMBER WITHIN A CERTAIN RANGE.

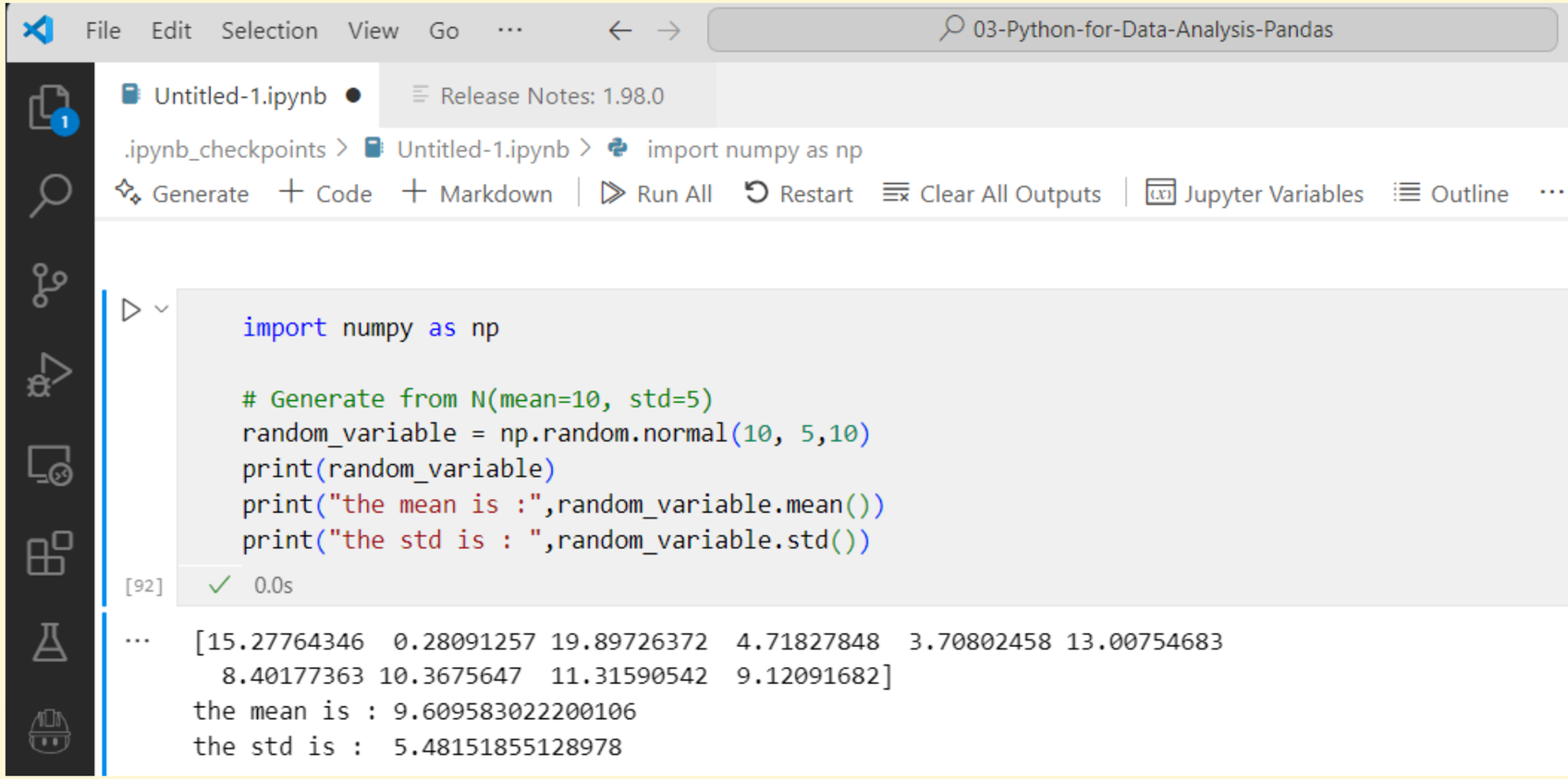
A . Uniform distribution

Each element has equal chance to be select.

CONTINUOUS RANDOM VARIABLE(CONT)

B . Normal distribution

Let’s generate 10 randoms normal variable with the mean of 10 and the standard deviation of 2 .

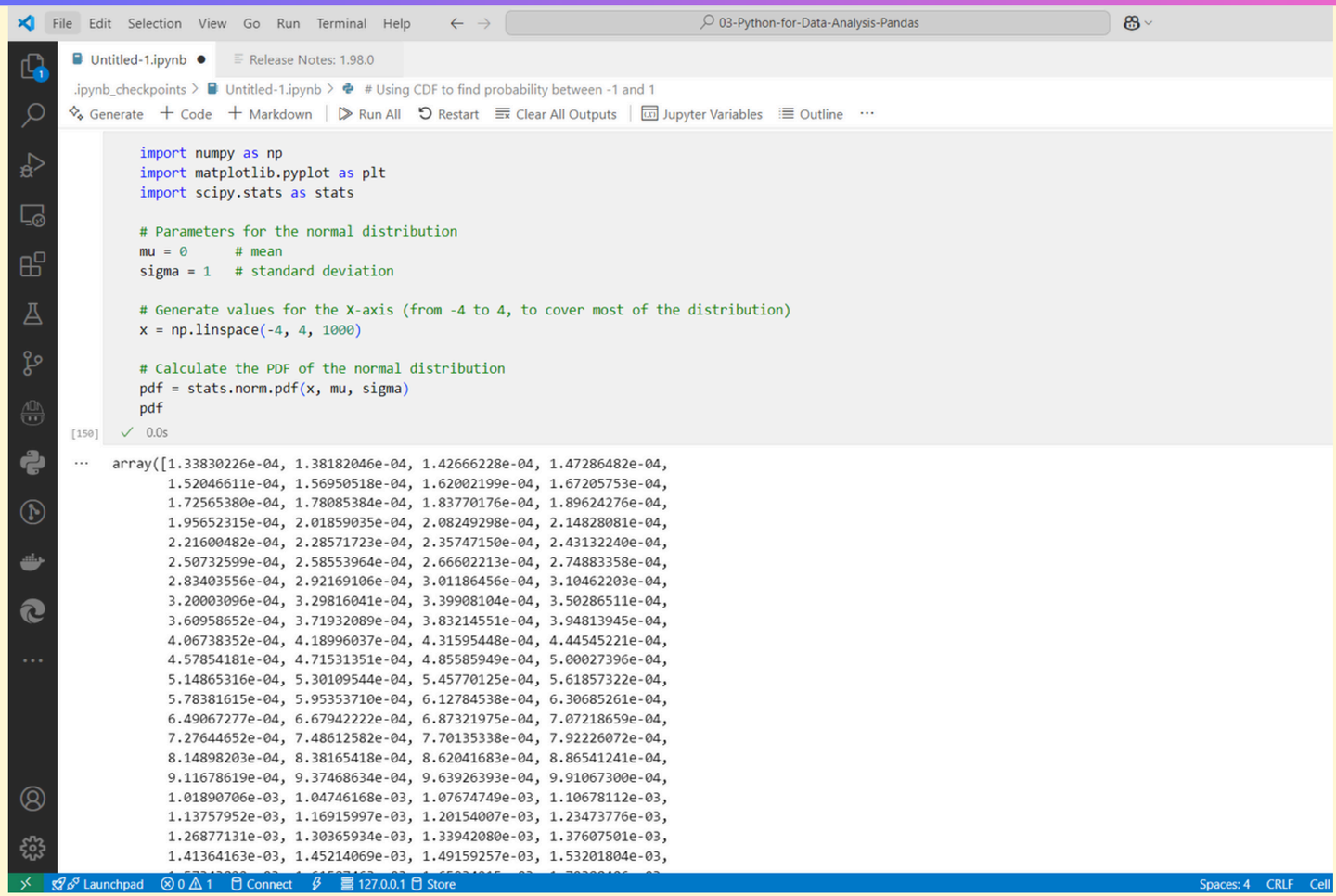


PDF CONTINUOUS RANDOM VARIABLE

Normal distribution

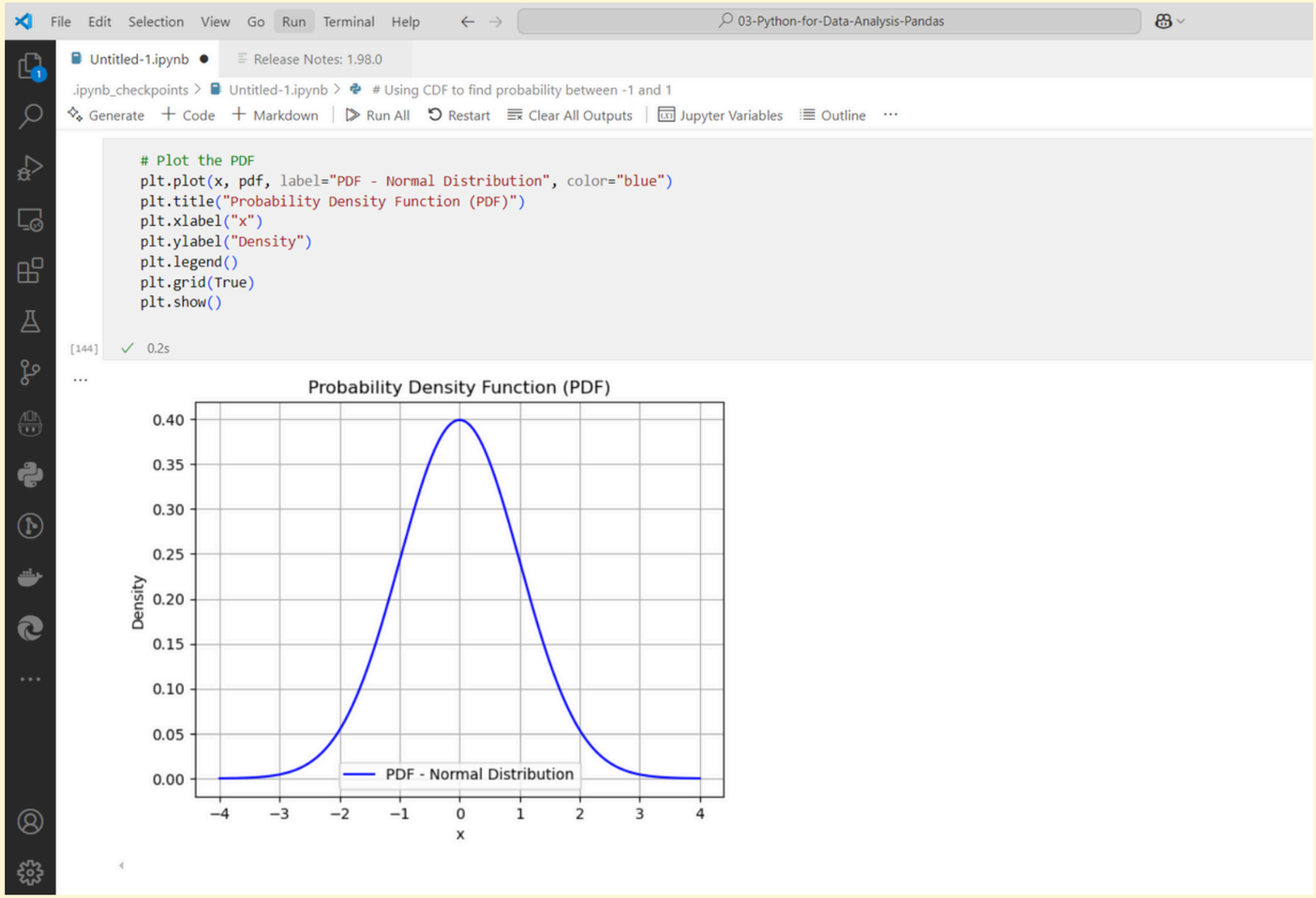
The Probability Density Function (PDF) is used for continuous random variables. It describes the likelihood of a random variable taking on a specific value.

The PDF gives a density, and the probability of the variable taking on a value in a specific range is calculated by integrating the PDF over that range.



PDF : THE BELL SHAPEP CURVE OF NORMAL DISTRIBUTION

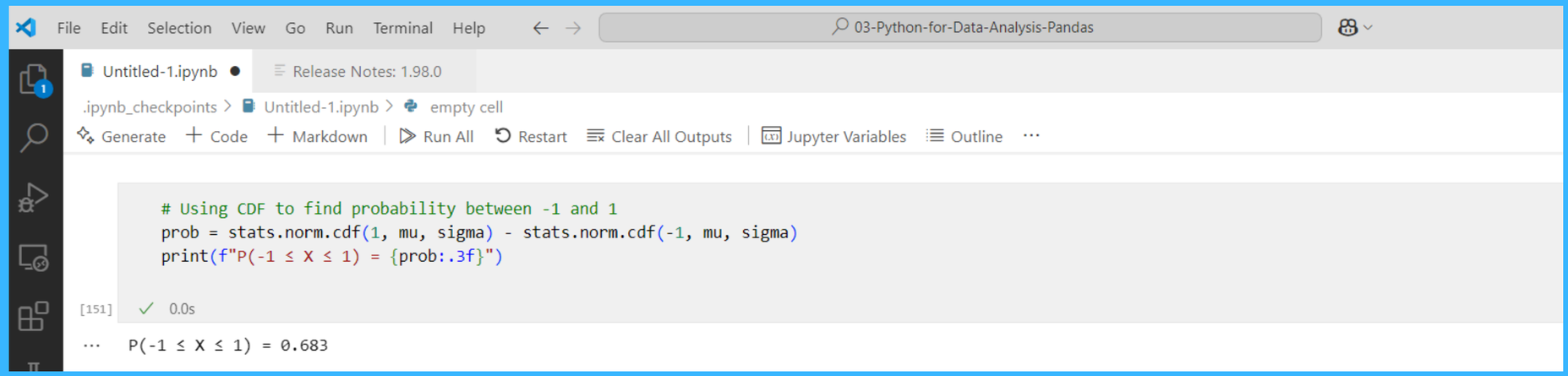
Normal distribution



CDF : FOR THE NORMAL DISTRIBUTION

PDF (Probability Density Function) is used for continuous random variables (probability at a specific value is zero, but you calculate probability over ranges).

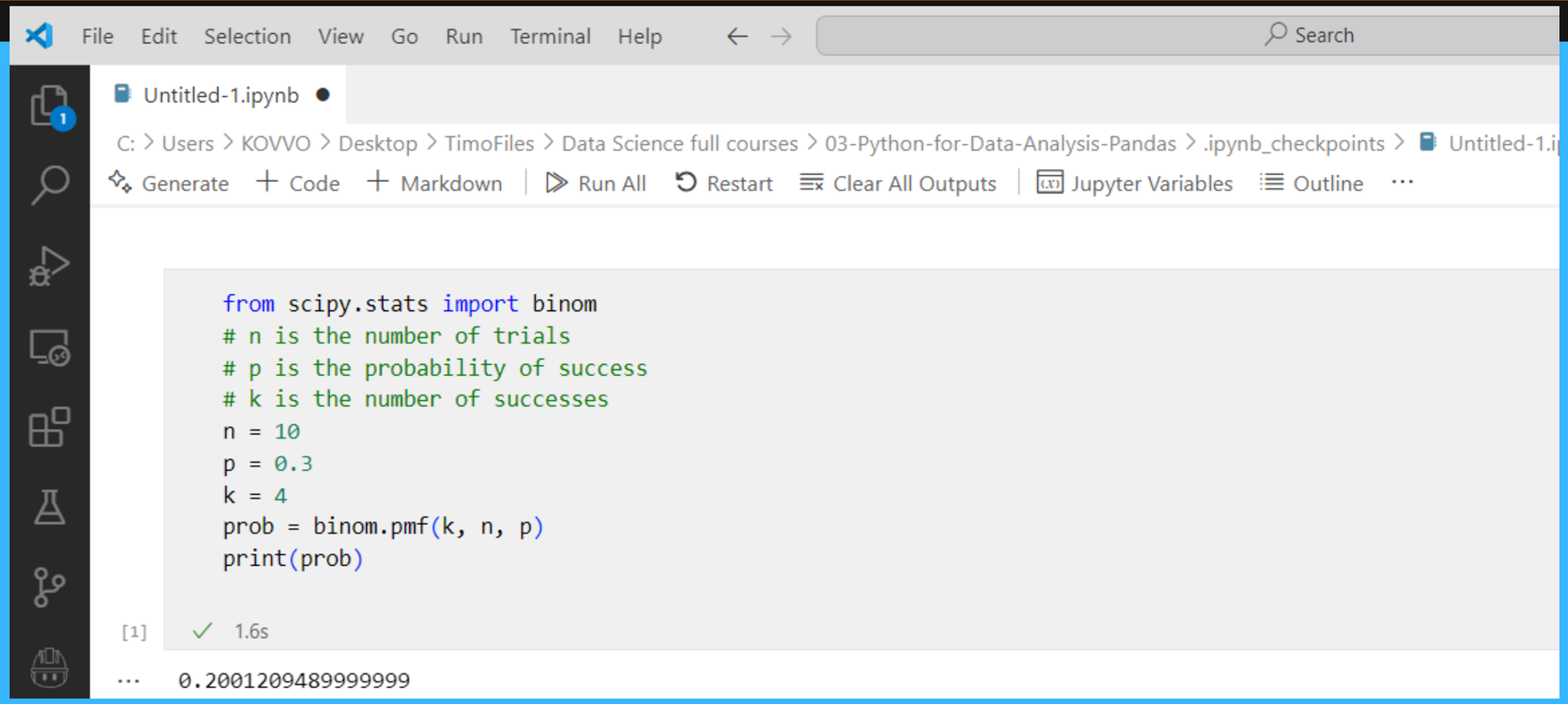
Let’s find the probability between -1 and 1.



BINOMIAL DISTRIBUTION

The binomial distribution is a discrete probability distribution that models the number of successes in a sequence of n independent trials, where each trial has a probability p of success.

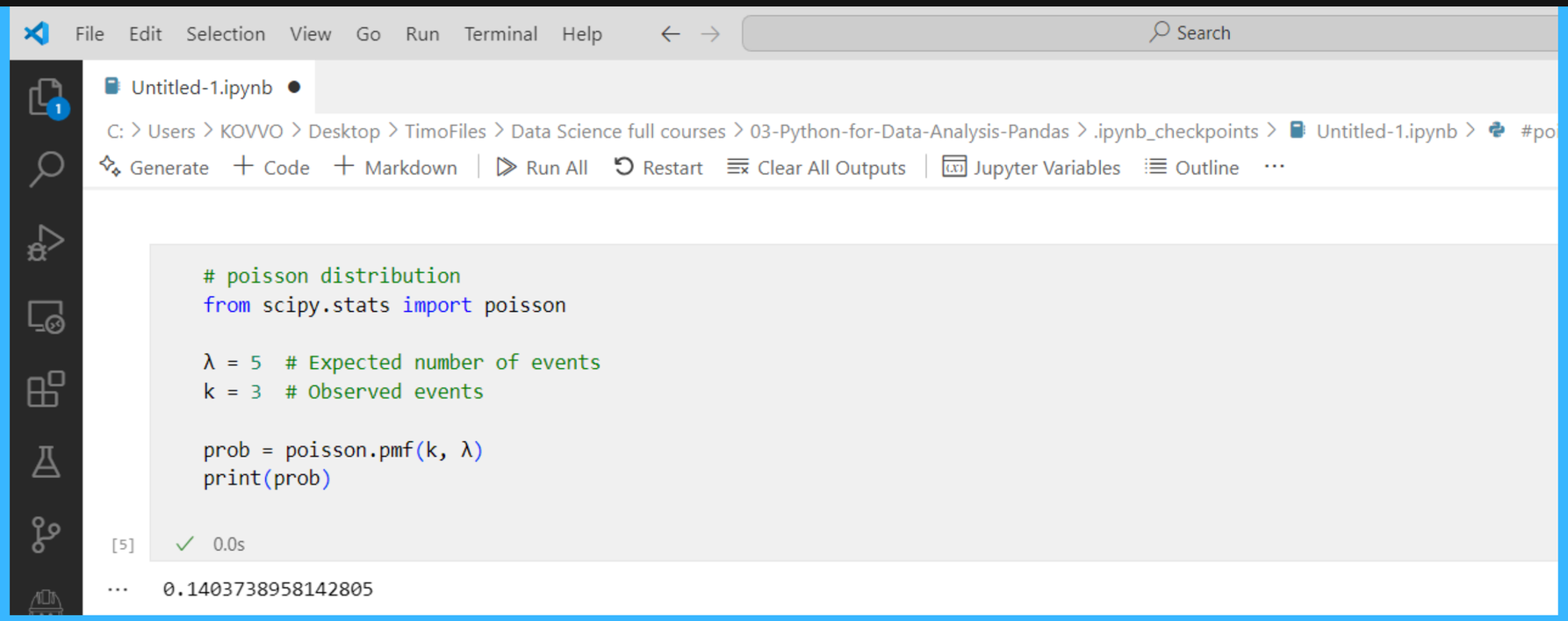
Suppose a test has a success probability of 0.3 and is repeated 10 times. The probability of getting exactly 4 successes is:



POISSON DISTRIBUTION

The Poisson distribution is a discrete probability distribution that models the number of events occurring in a fixed interval of time or space, given that these events happen independently and at a constant average rate λ.

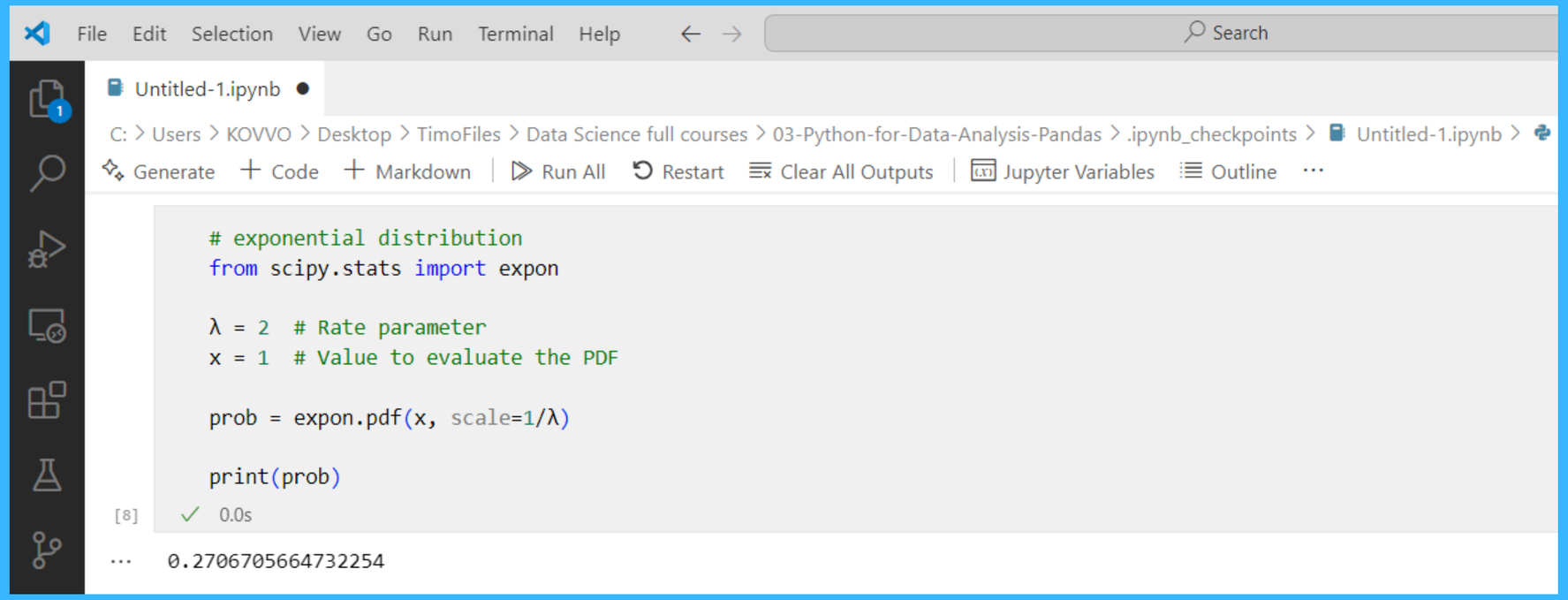
Suppose a website receives an average of 5 user sign-ups per hour. The probability of getting exactly 3 sign-ups in an hour is:



EXPONENTIAL DISTRIBUTION

The exponential distribution is a continuous probability distribution that models the time between independent events occurring at a constant average rate **λ**.

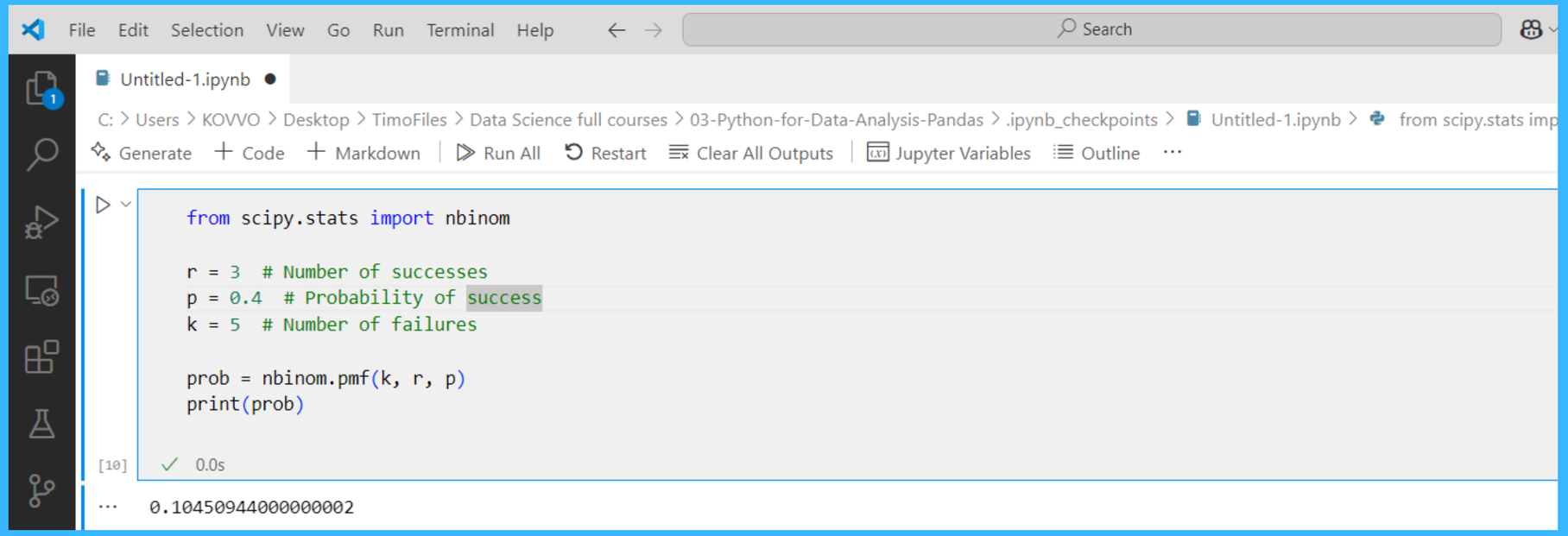
Suppose customer arrivals at a store follow an exponential distribution with an average arrival rate of **λ=2** customers per hour. The probability that the next customer arrives after more than 1 hour is:



NEGATIVE BINOMIAL DISTRIBUTION

The negative binomial distribution is a discrete probability distribution that models the number of trials required to achieve a fixed number of successes in a sequence of independent Bernoulli trials, where each trial has a probability p of success.

Suppose a basketball player has a 40% chance (p=0.4) of making a shot. What is the probability that they miss 5 times before making 3 successful shots?



THANK YOU