

Maestría en Ciencia de Datos (MCD)

Programación II



Unidad 1 Ciencia de datos con python

Programación II

Objetivo particular de la unidad: Conocer acerca de bibliotecas de relacionadas con el análisis de datos en Python para crear software de manera mas rápida y eficiente

1. Bibliotecas de ciencia de datos para Python

1.1 Ciencia de datos

1.2 Biblioteca NumPy

1.3 Biblioteca Pandas

1.4 Biblioteca rpy2

1.5 Biblioteca Scikit-learn

1.6 Distribución Anaconda

1.7 Visual Studio Code



Ciencia de Datos

Programación II

“La ciencia de datos es el estudio de datos con el fin de extraer información significativa para empresas. Es un enfoque multidisciplinario que combina principios y prácticas del campo de las matemáticas, la estadística, la inteligencia artificial y la ingeniería de computación para analizar grandes cantidades de datos.”
(AWS,2023)



Ciencia de Datos

Programación II

¿Por qué es importante la ciencia de datos?

La ciencia de datos es importante porque combina herramientas, métodos y tecnología para generar significado a partir de los datos.

(AWS,2023)



Ciencia de Datos

Programación II

Futuro de la ciencia de datos

La inteligencia artificial y las innovaciones del machine learning han hecho que el procesamiento de datos sea más rápido y eficiente. La demanda del sector ha creado un ecosistema de cursos, grados académicos y puestos de trabajo en el campo de la ciencia de datos.

(AWS,2023)



Ciencia de Datos

Programación II

¿Para qué se utiliza la ciencia de datos?

La ciencia de datos se utiliza para estudiar los datos de cuatro maneras principales:

1. Análisis descriptivo

El análisis descriptivo examina los datos para obtener información sobre lo que ha ocurrido u ocurre en el entorno de datos. Se caracteriza por las visualizaciones de datos, como los gráficos circulares, de barras o líneas, las tablas o las narraciones generadas.

(AWS,2023)



Ciencia de Datos

Programación II

2. Análisis de diagnóstico

El análisis de diagnóstico es un examen profundo o detallado de datos para entender por qué ha ocurrido algo. Se caracteriza por técnicas como el análisis detallado, el descubrimiento y la minería de datos o las correlaciones. Se pueden llevar a cabo varias operaciones y transformaciones de datos en un determinado conjunto con el fin de descubrir patrones únicos en cada una de estas técnicas.

(AWS,2023)



Ciencia de Datos

Programación II

3. Análisis predictivo

El análisis predictivo utiliza los datos históricos para hacer previsiones precisas sobre los patrones de datos que pueden producirse en el futuro. Se caracteriza por técnicas como el machine learning, la previsión, la coincidencia de patrones y el modelado predictivo

(AWS,2023)



Ciencia de Datos

Programación II

3. Análisis predictivo

El análisis predictivo utiliza los datos históricos para hacer previsiones precisas sobre los patrones de datos que pueden producirse en el futuro. Se caracteriza por técnicas como el machine learning, la previsión, la coincidencia de patrones y el modelado predictivo

(AWS,2023)



Ciencia de Datos

Programación II

4. Análisis prescriptivo

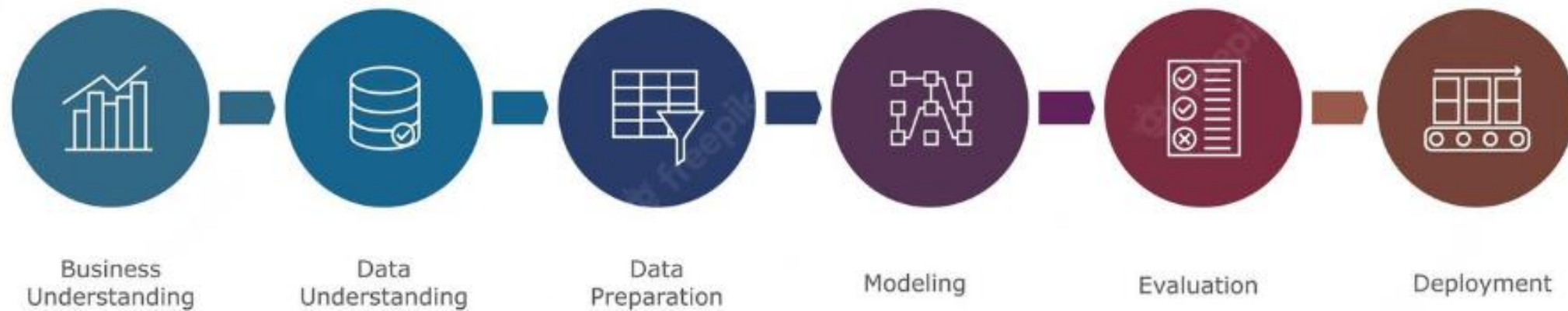
El análisis prescriptivo lleva los datos predictivos al siguiente nivel. No solo predice lo que es probable que ocurra, sino que sugiere una respuesta óptima para ese resultado. Puede analizar las posibles implicaciones de las diferentes alternativas y recomendar el mejor curso de acción.

(AWS,2023)



Ciencia de Datos

Programación II



Anaconda

Programación II



¿Qué es el Navegador Anaconda?

Anaconda Navigator es una interfaz gráfica de usuario (GUI) de escritorio incluida en la distribución de Anaconda® que le permite iniciar aplicaciones y administrar fácilmente paquetes, entornos y canales de conda sin usar comandos de línea de comandos. Navigator puede buscar paquetes en Anaconda.org o en un repositorio local de Anaconda. Está disponible para Windows, macOS y Linux.

(Anaconda, 2023)



Anaconda

Programación II

¿A qué aplicaciones puedo acceder usando Navigator?

Las siguientes aplicaciones están disponibles de forma predeterminada en Navigator:

- [JupyterLab](#)
- [Cuaderno Jupyter](#)
- [espía](#)
- [PyCharm](#)
- [Glueviz](#)
- [Aplicación naranja 3](#)
- [RStudio](#)
- Aviso de Anaconda (solo Windows)
- Anaconda PowerShell (solo Windows)



Anaconda

Programación II

Data science technology for a better world.

Anaconda offers the easiest way to perform Python/R data science and machine learning on a single machine. Start working with thousands of open-source packages and libraries today.

Download 

For Windows

Python 3.9 • 64-Bit Graphical Installer • 621 MB

Get Additional Installers



<https://www.anaconda.com/>





















Visual Studio Code

Programación II

ANACONDA.NAVIGATOR Upgrade Now Connect

Home | Environments | Learning | Community

All applications on base (root) Channels

 <p>DataSpell</p> <p>DataSpell is an IDE for exploratory data analysis and prototyping machine learning models. It combines the interactivity of Jupyter notebooks with the intelligent Python and R coding assistance of PyCharm in one user-friendly environment.</p> <p>Install</p>	 <p>CMD.exe Prompt 0.1.1</p> <p>Run a cmd.exe terminal with your current environment from Navigator activated</p> <p>Launch</p>	 <p>JupyterLab 3.5.2</p> <p>An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.</p> <p>Launch</p>	 <p>Jupyter Notebook 6.5.2</p> <p>Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.</p> <p>Launch</p>	 <p>Powershell Prompt 0.0.1</p> <p>Run a Powershell terminal with your current environment from Navigator activated</p> <p>Launch</p>	 <p>Qt Console 5.2.2</p> <p>PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.</p> <p>Launch</p>
 <p>Spyder 5.2.2</p> <p>Scientific Python Development Environment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features</p> <p>Launch</p>	 <p>VS Code 1.74.3</p> <p>Streamlined code editor with support for development operations like debugging, task running and version control.</p> <p>Launch</p>	 <p>Datalore</p> <p>Kick-start your data science projects in seconds in a pre-configured environment. Enjoy coding assistance for Python, SQL, and R in Jupyter notebooks and benefit from no-code automations. Use Datalore online for free.</p> <p>Launch</p>	 <p>Deepnote</p> <p>Deepnote is a new kind of data notebook build for collaboration - Jupyter compatible, in the cloud and sharing is easy as sending a link</p> <p>Launch</p>	 <p>IBM Watson Studio Cloud</p> <p>IBM Watson Studio Cloud provides you the tools to analyze and visualize data, to cleanse and shape data, to create and train machine learning models. Prepare data and build models, using open source data science tools or visual modeling.</p> <p>Launch</p>	 <p>ORACLE Cloud Infrastructure Oracle Data Science Service</p> <p>OCI Data Science offers a machine learning platform to build, train, manage, and deploy your machine learning models on the cloud with your favorite open-source tools</p> <p>Launch</p>
 <p>console_shortcut_miniconda 0.1.1</p>	 <p>Glueviz 1.2.4</p> <p>Multidimensional data visualization across files. Explore relationships within and</p>	 <p>Orange 3 3.32.0</p> <p>Component based data mining framework. Data visualization and data analysis for</p>	 <p>powershell_shortcut_miniconda 0.0.1</p>	 <p>PyCharm Professional</p> <p>A Full-Fledged IDE by JetBrains for both Scientific and Web Python development.</p>	 <p>RStudio 1.1.456</p> <p>A set of integrated tools designed to help you be more productive with R. Includes R</p>

Anaconda Notebooks: Cloud notebooks with hundreds of packages ready to code. [Learn More](#)

Documentation | Anaconda Blog

Twitter | YouTube | GitHub



Visual Studio Code

Programación II



Visual Studio Code

Qué es Visual Studio Code?

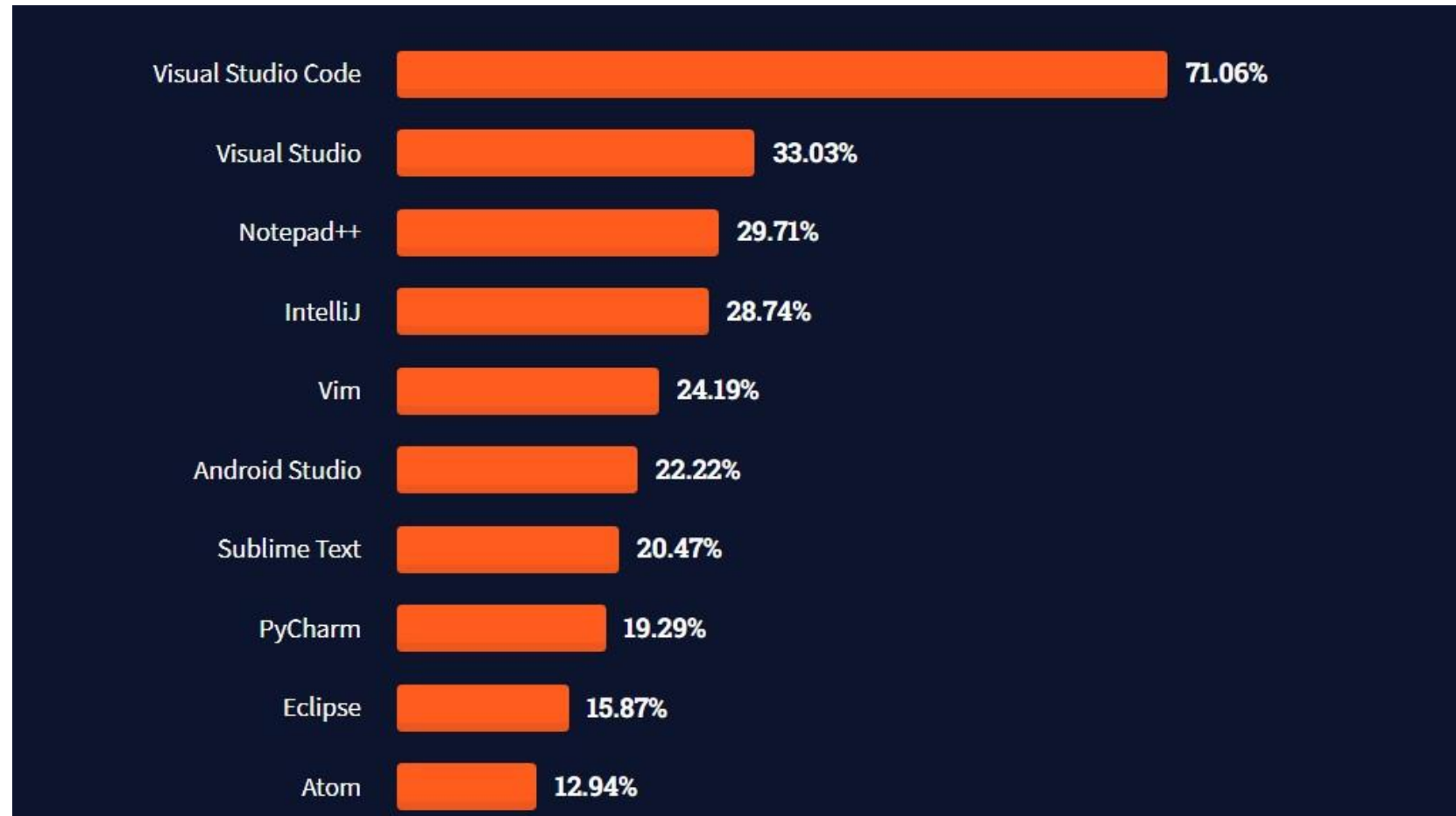
Visual Studio Code (VS Code) es un editor de código fuente desarrollado por Microsoft. Es software libre y multiplataforma, está disponible para Windows, GNU/Linux y macOS. VS Code tiene una buena integración con Git, cuenta con soporte para depuración de código, y dispone de un sinnúmero de extensiones, que básicamente te da la posibilidad de escribir y ejecutar código en cualquier lenguaje de programación

(OpenWebinars, 2023)



Visual Studio Code

Programación II

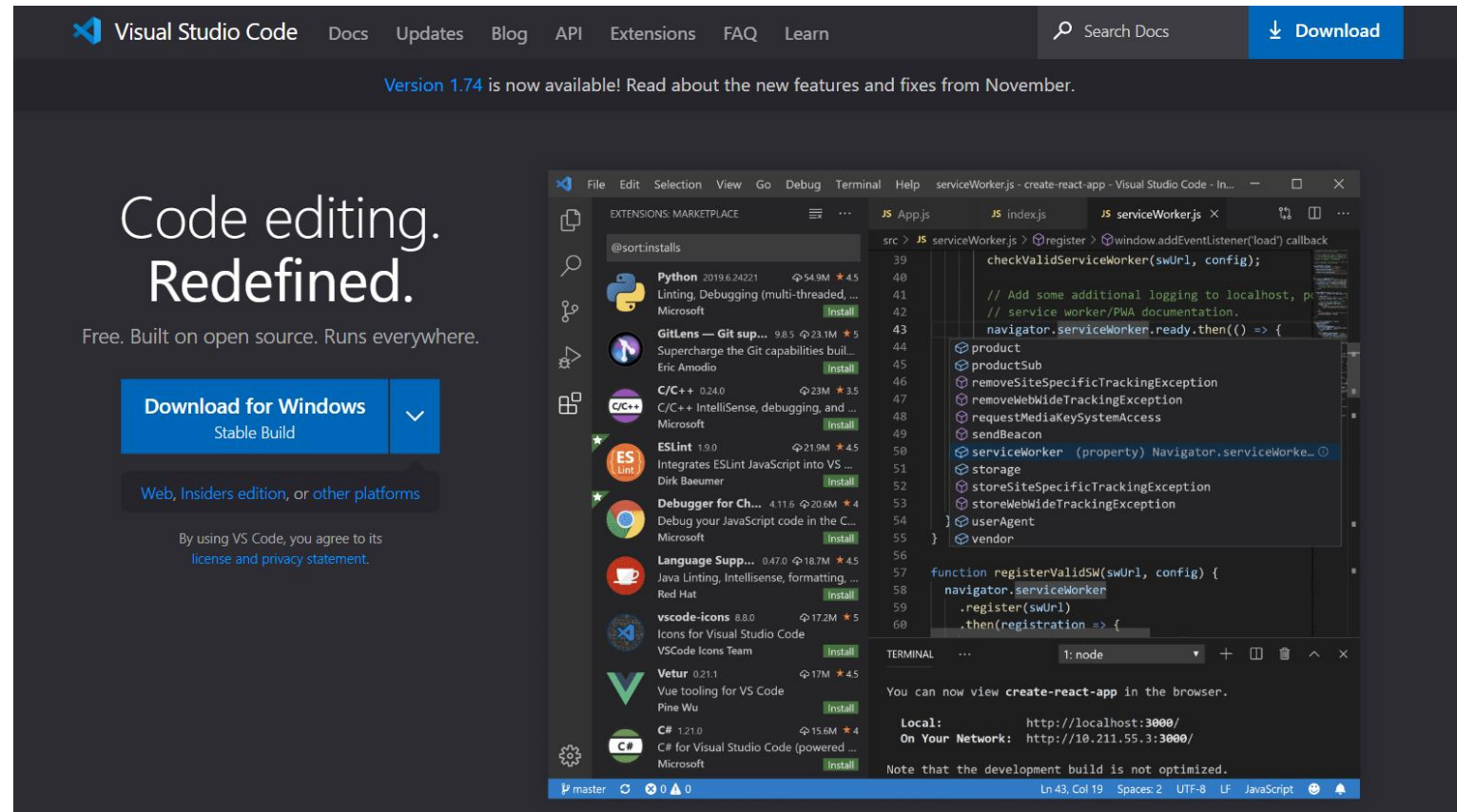


(OpenWebinars, 2023)



Visual Studio Code

Programación II



<https://code.visualstudio.com/>

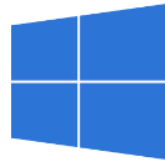


Visual Studio Code

Programación II

Download Visual Studio Code

Free and built on open source. Integrated Git, debugging and extensions.



↓ Windows

Windows 8, 10, 11

User Installer	x64	x86	Arm64
System Installer	x64	x86	Arm64
.zip	x64	x86	Arm64
CLI	x64	x86	Arm64



↓ .deb

Debian, Ubuntu

↓ .rpm

Red Hat, Fedora, SUSE

.deb	x64	Arm32	Arm64
.rpm	x64	Arm32	Arm64
.tar.gz	x64	Arm32	Arm64
Snap	Snap Store		
CLI	x64	Arm32	Arm64



↓ Mac

macOS 10.11+

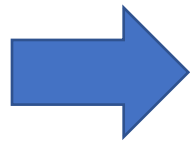
.zip	Intel chip	Apple silicon	Universal
CLI	Intel chip	Apple silicon	

<https://code.visualstudio.com/>



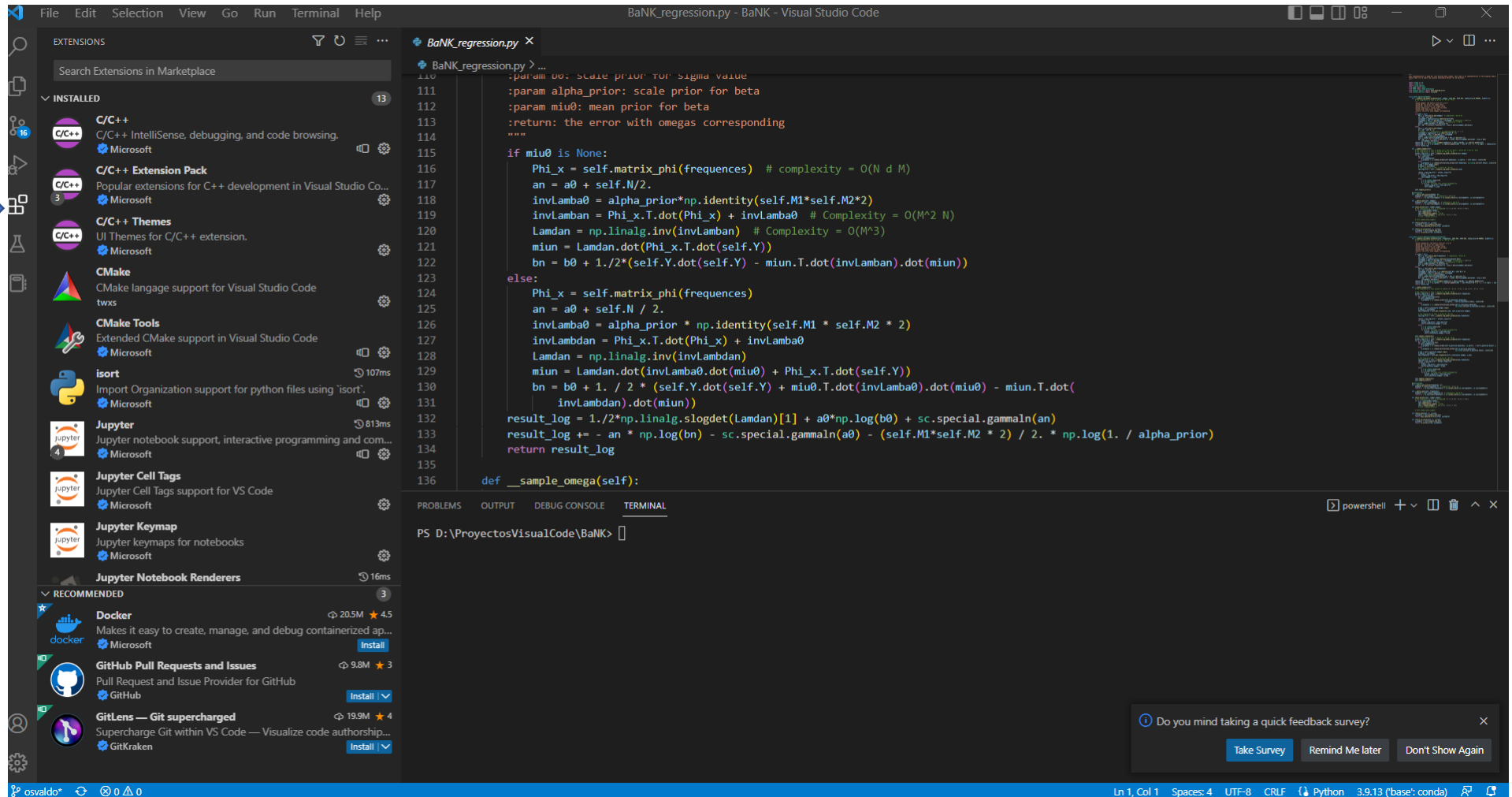
Visual Studio Code

Programación II



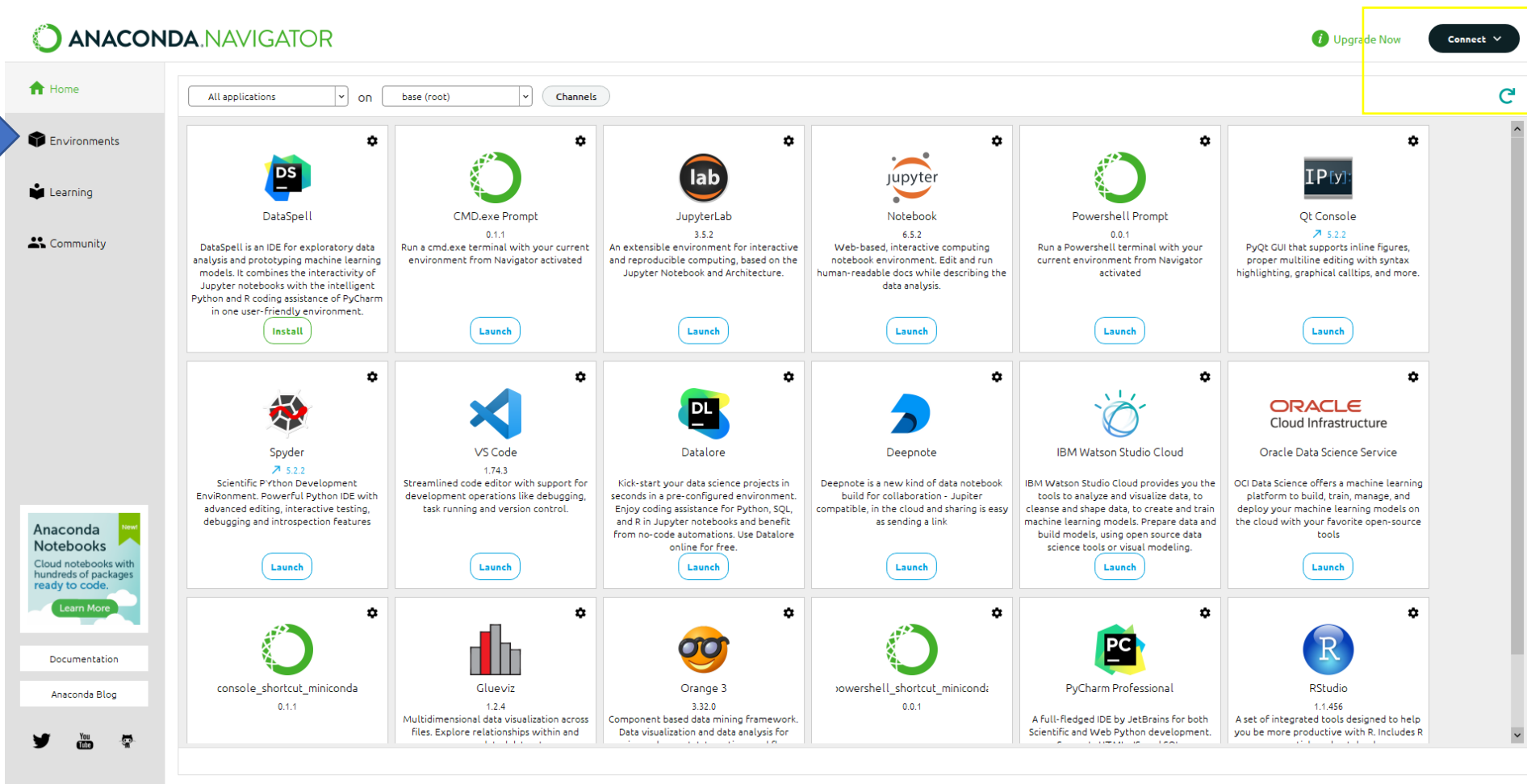
Actividades

- Instalar Python
- Instalar Jupyter
- Crear carpeta
- Abrir carpeta



Environments

Programación II



ANACONDA.NAVIGATOR

Home
Environments
Learning
Community

All applications on base (root) Channels

DataSpell
0.1.1
DataSpell is an IDE for exploratory data analysis and prototyping machine learning models. It combines the interactivity of Jupyter notebooks with the intelligent Python and R coding assistance of PyCharm in one user-friendly environment.
[Install](#)

CMD.exe Prompt
0.1.1
Run a cmd.exe terminal with your current environment from Navigator activated
[Launch](#)

JupyterLab
3.5.2
An extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.
[Launch](#)

Jupyter Notebook
6.5.2
Web-based, interactive computing notebook environment. Edit and run human-readable docs while describing the data analysis.
[Launch](#)

Powershell Prompt
0.0.1
Run a Powershell terminal with your current environment from Navigator activated
[Launch](#)

Qt Console
5.2.2
PyQt GUI that supports inline figures, proper multiline editing with syntax highlighting, graphical calltips, and more.
[Launch](#)

Spyder
5.2.2
Scientific Python Development Environment. Powerful Python IDE with advanced editing, interactive testing, debugging and introspection features
[Launch](#)

VS Code
1.74.3
Streamlined code editor with support for development operations like debugging, task running and version control.
[Launch](#)

Datalore
Kick-start your data science projects in seconds in a pre-configured environment. Enjoy coding assistance for Python, SQL, and R in Jupyter notebooks and benefit from no-code automations. Use Datalore online for free.
[Launch](#)

Deepnote
Deepnote is a new kind of data notebook build for collaboration - Jupyter compatible, in the cloud and sharing is easy as sending a link
[Launch](#)

IBM Watson Studio Cloud
IBM Watson Studio Cloud provides you the tools to analyze and visualize data, to cleanse and shape data, to create and train machine learning models. Prepare data and build models, using open source data science tools or visual modeling.
[Launch](#)

ORACLE Cloud Infrastructure
Oracle Data Science Service
OCI Data Science offers a machine learning platform to build, train, manage, and deploy your machine learning models on the cloud with your favorite open-source tools
[Launch](#)

console_shortcut_miniconda
0.1.1
[Launch](#)

Glueviz
1.2.4
Multidimensional data visualization across files. Explore relationships within and
[Launch](#)

Orange 3
3.32.0
Component based data mining framework. Data visualization and data analysis for
[Launch](#)

powershell_shortcut_miniconda
0.0.1
[Launch](#)




PyCharm Professional
A Full-Featured IDE by JetBrains for both Scientific and Web Python development.
[Launch](#)

RStudio
1.1.456
A set of integrated tools designed to help you be more productive with R. Includes R
[Launch](#)

[Upgrade Now](#) [Connect](#)

Anaconda Notebooks
Cloud notebooks with hundreds of packages ready to code.
[Learn More](#)

Documentation
Anaconda Blog

Environments

Programación II

Environments

File Help

ANACONDA.NAVIGATOR

Upgrade Now Connect

Home Environments Learning Community

Search Environments

base (root) bank computervision gwp gwp_tensor gwp_tf_mflow latam medicalimages mlflow objectdetec

Installed Channels Update index...

Search Packages

Dependencies installed

Name	Description	Version
✓ _ipyw_jlab_nb_ex...		0.1.0
✓ alabaster		0.7.12
✓ anaconda		2022.10
✓ anaconda-client		1.11.0
✓ anaconda-project		0.11.1
✓ anyio		3.5.0
✓ appdirs		1.4.4
✓ argon2-cffi		21.3.0
✓ argon2-cffi-bindings		21.2.0
✓ arrow		1.2.3
✓ astroid		2.11.7
✓ astropy		5.1
✓ atomicwrites		1.4.0
✓ attrs		22.1.0
✓ automat		20.2.0
✓ autopep8		1.6.0
✓ babel		2.11.0
✓ backcall		0.2.0

Create Clone Import Backup Remove

Anaconda Notebooks
Cloud notebooks with hundreds of packages ready to code.
Learn More

Documentation
Anaconda Blog

Twitter YouTube GitHub

Environments

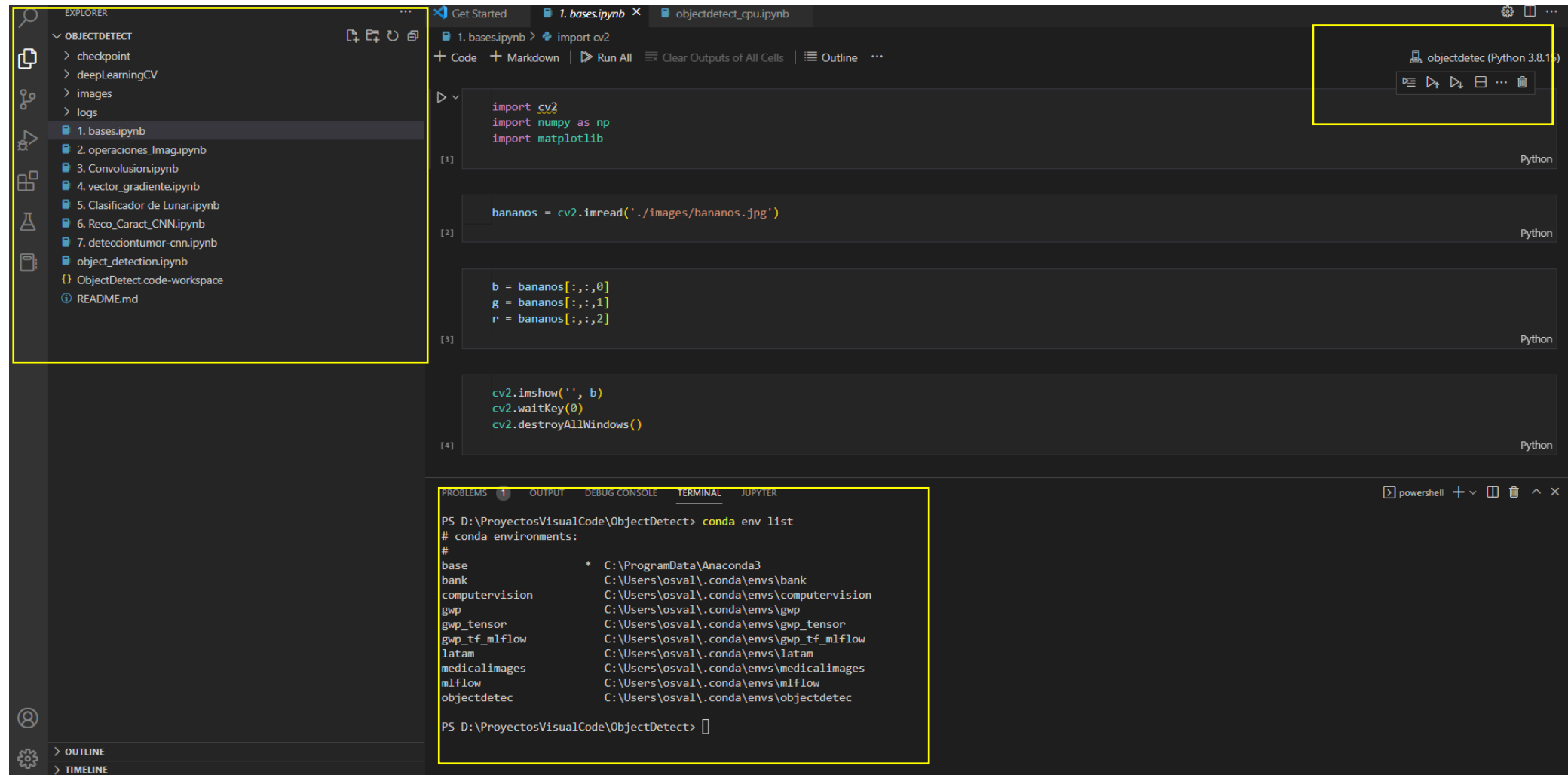
Programación II

- *conda create -n nombreenv python=3.x*
- *conda remove -n nombreenv --all*
- *conda env list*
- *conda activate nombreenv*
- *conda deactivate*



Environments (Problema común)

Programación II



The screenshot displays a JupyterLab environment. On the left, the Explorer panel shows a project named 'OBJECTDETECT' with the following structure:

- checkpoint
- deepLearningCV
- images
- logs
- 1. bases.ipynb
- 2. operaciones_Imag.ipynb
- 3. Convolution.ipynb
- 4. vector_gradiente.ipynb
- 5. Clasificador de Lunar.ipynb
- 6. Reco_Caract_CNN.ipynb
- 7. detecciontum-cnn.ipynb
- object_detection.ipynb
- ObjectDetect.code-workspace
- README.md

The main editor shows the '1. bases.ipynb' file with the following code:

```
import cv2
import numpy as np
import matplotlib

bananos = cv2.imread('./images/bananos.jpg')

b = bananos[:, :, 0]
g = bananos[:, :, 1]
r = bananos[:, :, 2]

cv2.imshow('', b)
cv2.waitKey(0)
cv2.destroyAllWindows()
```

The terminal at the bottom shows the output of the command 'conda env list':

```
PS D:\ProyectosVisualCode\ObjectDetect> conda env list
# conda environments:
#
base                  *  C:\ProgramData\Anaconda3
bank                  C:\Users\osval\.conda\envs\bank
computervision        C:\Users\osval\.conda\envs\computervision
gwp                   C:\Users\osval\.conda\envs\gwp
gwp_tensor             C:\Users\osval\.conda\envs\gwp_tensor
gwp_tf_mflow           C:\Users\osval\.conda\envs\gwp_tf_mflow
latam                  C:\Users\osval\.conda\envs\latam
medicalimages          C:\Users\osval\.conda\envs\medicalimages
mlflow                 C:\Users\osval\.conda\envs\mlflow
objectdetec            C:\Users\osval\.conda\envs\objectdetec

PS D:\ProyectosVisualCode\ObjectDetect>
```


Environments (Solución)

Programación II

