

Mi vida después del bootcamp

Data Science Bootcamp

The Bridge



Ya he terminado el bootcamp...

¿Y ahora qué?



I SHOULD START A JOB SEARCH



WE ARE READY



FOR THAT VACATION

Tiempo de reset y pensar por
dónde tirar

Vale, ¿y después qué?

¿Qué te gusta/se te da bien?

De todo lo que has visto, ¿con qué te quedas?

Ramp Up

1. Programación con Python
2. Programación orientada a objetos

Data Analysis

1. Acceso a datos: APIs, Bases de datos, SQL
2. Limpieza de datos: Numpy, Pandas
3. Automatización: de procesos, web scrapping
4. Visualización con Python
5. Dashboards de Streamlit
6. Analítica de los datos
7. Estadística
8. Business Intelligence
9. Mapas

Machine Learning

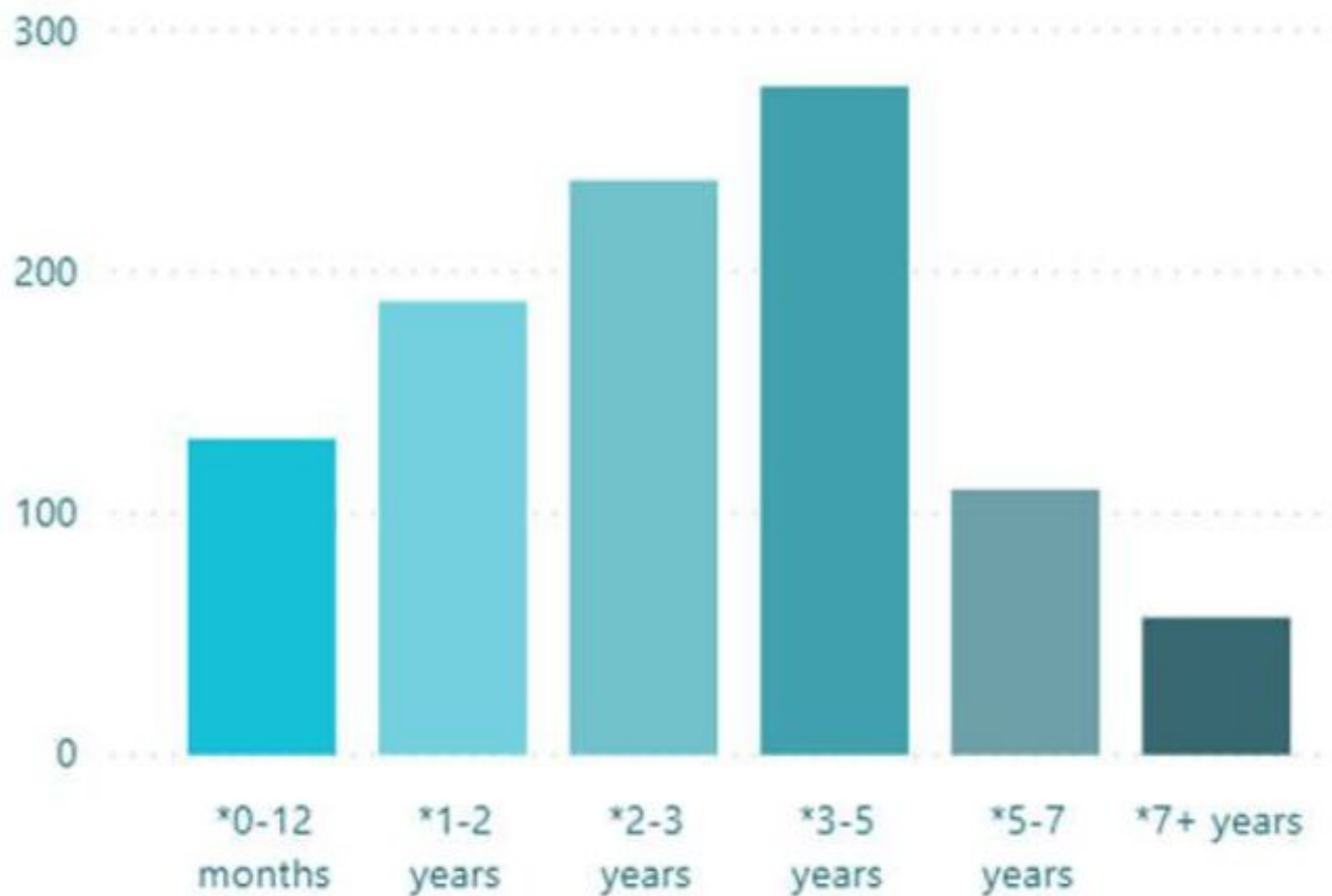
1. Automatizar toma de decisiones
2. Algoritmo supervisados
3. Comprensión de los algoritmos y variables que impactan en la toma de decisión
4. Algoritmos no supervisados
5. NLP
6. Deep Learning
7. Time Series
8. Aprendizaje por refuerzo

Data Engineer

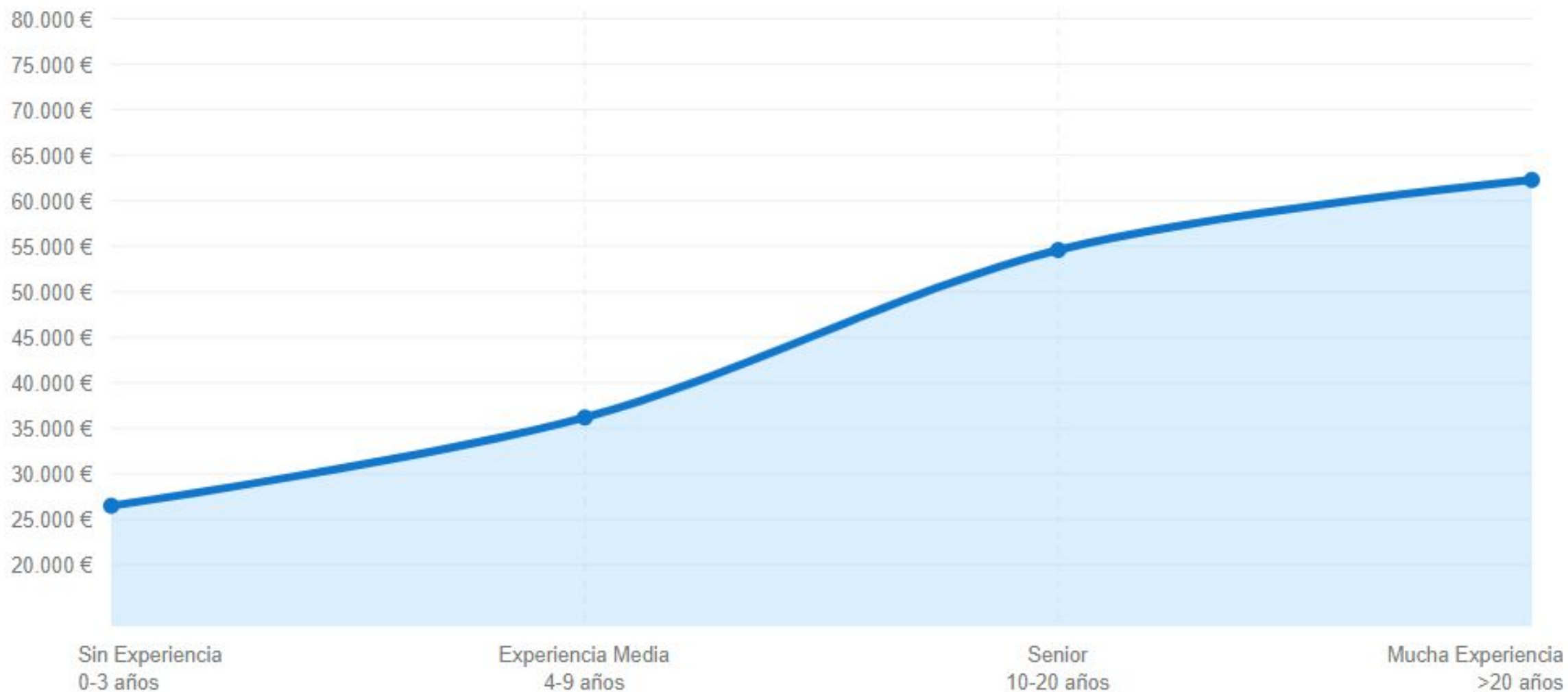
1. APIs
2. Cloud, despliegue de recursos
3. Máquinas, servidores, redes
4. Big Data & Spark
5. Airflow
6. Docker

¿Cómo está el mercado?

Years on the Job



Cuánto Gana un Data Scientist según de la Experiencia



Las entrevistas de Data Engineer aumentaron un 40% en 2020

Las entrevistas de Data Analysis aumentaron un 20% en 2020

En 2020 las entrevistas de Data Science han subido un 10% cuando en años anteriores era de 80%

La mayoría de preguntas de entrevista fueron de Machine Learning, SQL, código, algoritmia y estadística.

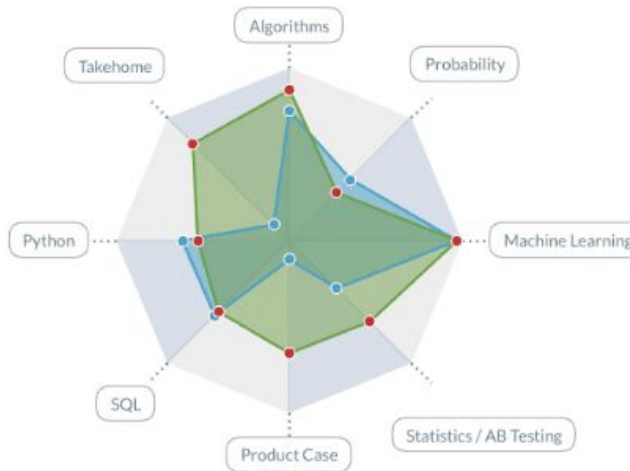
Mandaron ejercicios en un 25% de las entrevistas

[Basado en este informe](#)

¿Qué tipos de preguntas se hacen en una entrevista técnica?

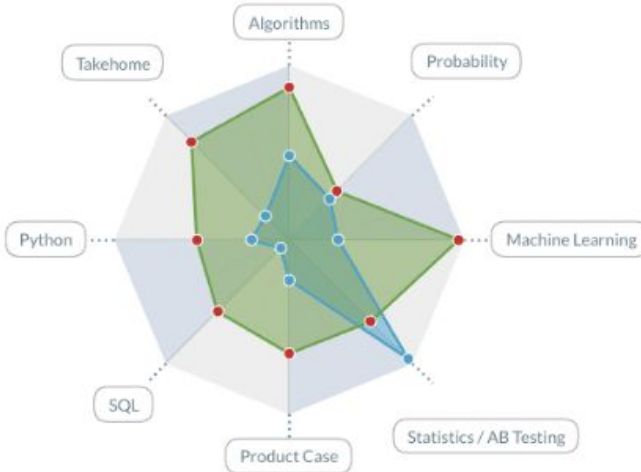
DATA SCIENTIST INTERVIEWS

amazon



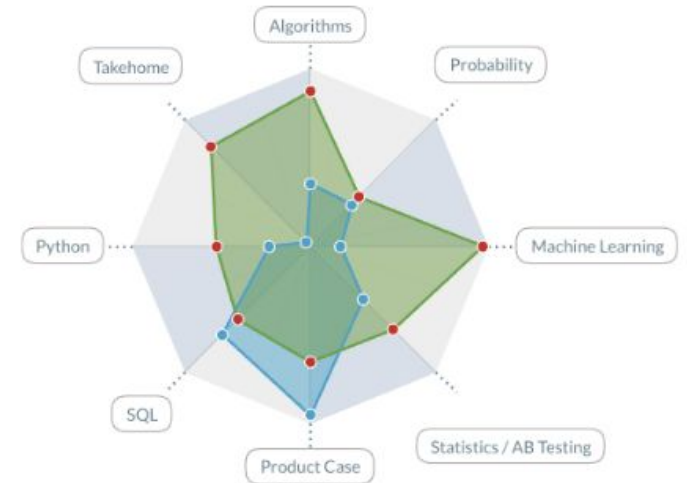
Amazon Data Scientist Average Data Scientist interview
Powered by ZingChart

Google



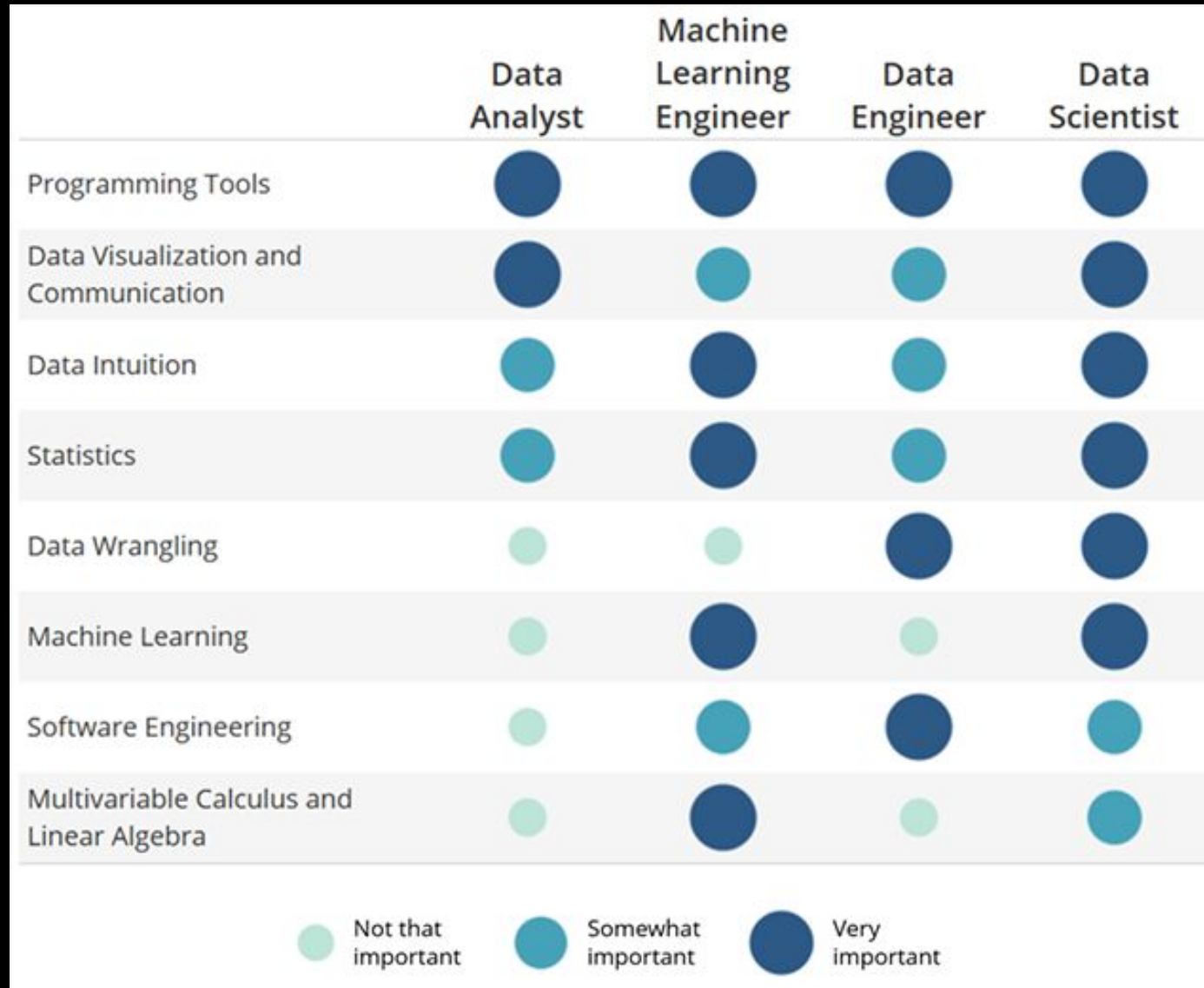
Google Data Scientist Average Data Scientist interview
Powered by ZingChart

facebook



Facebook Data Scientist Average Data Scientist interview
Powered by ZingChart

Puestos de trabajo de Data Science



Data Analyst

SKILLS

1. **Analítica**: exploratorio, obtener insights, encontrar patrones
2. **Acceso a Bases de datos**: interno y externo a la empresa.
3. **Data Cleaning**
4. **Data Quality**
5. **Business understanding**
6. **Communication**: presentar resultados. Storytelling.
7. **Mates y estadística**
8. **Project manager**
9. **Dashboards/Reportes**: Automatizado
10. **Machine Learning**: no experto

TOOLS

1. **Lenguajes de programación**: SQL, Python, R
2. **Business intelligence**: Power BI, Tableau, Microstrategy, QlikSense, DataStudio.
3. **Relational DB**: MySQL, SQLServer, Oracle, MariaDB.
4. **Non relational DB**: Cassandra, MongoDB
5. **Excel**
6. **IDEs**: Pycharm, Anaconda, Jupyter, Rstudio
7. **Big Data**: pyspark Hive, ecosistema Cloudera

Machine Learning Engineer



SKILLS

1. **Modelos**: Regresión lineal, SVM, Kmeans...
2. **Matemáticas y estadística**: álgebra, probabilidad, estadística, cálculo, distribuciones...
3. **Tratamiento de imágenes**
4. **Redes neuronales**
5. **NLP**
6. **Comunicación**
7. **Conocimiento de negocio**
8. **Avanzado**: robótica, vídeo, integración con sistemas, automatización

TOOLS

1. **Lenguajes de programación**: SQL, Python, R, C++ (comp. paralela)
2. **Deep Learning**: Tensorflow, Keras, Pytorch
3. **Relational DB**: MySQL, SQLServer, Oracle, MariaDB.
4. **Non relational DB**: Cassandra, MongoDB
5. **IDEs**: Pycharm, Anaconda, Jupyter, Rstudio
6. **Big Data**: pyspark Hive, Spark MLlib, ecosistema Cloudera
7. **Cloud**: soluciones de IA en cloud.
8. **Automation**: MLOps
9. **Otros**: MLFlow, Tensorboard

Data Engineer

Responsibilities...

ON DAILY BASIS

Develop, construct, test, and maintain architectures (such as databases and large-scale processing systems)



Ensure architecture will support the requirements of the business



Discover opportunities for data acquisition



Develop data set processes for data modeling, mining and production



Employ a variety of languages and tools (e.g. scripting languages) to marry systems together



Recommend ways to improve data reliability, efficiency and quality



Conduct research to answer industry and business questions



Leverage large volumes of data from internal and external sources to answer that business



Employ sophisticated analytics programs, machine learning and statistical methods to prepare data for use in predictive and prescriptive modeling



Explore and examine data to find hidden patterns

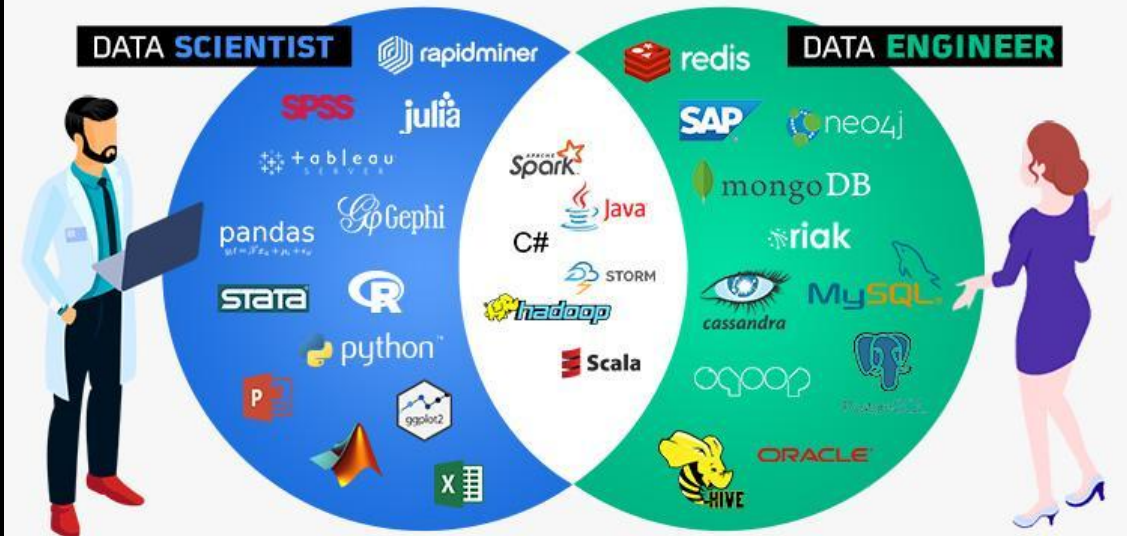


Automate work through the use of predictive and prescriptive analytics



Tell stories to key stakeholders based on their analysis

LANGUAGES, TOOLS AND SOFTWARE



Data Engineering Tools



Next steps Data Engineer

1. **Bases de datos**: montar y gestionar bases de datos relacionales y no relacionales.
2. **Data Warehouse**: ETLs, ingestas de datos, modelado de datos para su posterior consumo.
3. **Redes**: comunicaciones, infraestructura, acceso a máquinas, bash, scripting
4. **Cloud**: IaaS, PaaS, seguridad, acceso usuarios, despliegues automáticos, migraciones a cloud.
5. **Virtualización y contenedores**: Docker y Kubernetes. Arquitecturas de microservicios.
6. **Big Data**: ecosistema Hadoop, Spark, Spark Streaming, Databricks, integración con Cloud.
7. **Automatización con pipelines**: scripting con Python, Airflow, Luigi, crontab.
8. **DevOps**: Automatización de puestas en producción.
9. **Real time**: Kafka, Spark Streaming, soluciones cloud.
10. **Otros lenguajes**: Python, SQL, Java, Scala, C, Julia.

Otros roles

DATA ARCHITECT

THE CONTEMPORARY DATA MODELLER

Languages

SQL, XML, Hive, Pig, Spark

Skills & Talents

- ✓ Data warehousing solutions
- ✓ In-depth knowledge of database architecture
- ✓ Extraction Transformation and Load (ETL), spreadsheet and BI tools
- ✓ Data modeling
- ✓ Systems development



Role:

Creates blueprints for data management systems to integrate, centralize, protect and maintain data sources

Mindset:

Inquiring ninja with a love for data architecture design patterns

HIRED BY



STATISTICIAN

'HISTORIC LEADERS OF DATA'

Languages

R, SAS, SPSS, Matlab, Stata, Python, Perl, Hive, Pig, Spark, SQL

Skills & Talents

- ✓ Statistical theories & methodology
- ✓ Data mining & machine learning
- ✓ Distributed Computing (Hadoop)
- ✓ Database systems (SQL and NO SQL based)
- ✓ Cloud tools



Role

Collects, analyzes and interprets-qualitative as well as quantitative data with statistical theories and methods

Mindset

Logical and enthusiastic stats genius

HIRED BY



DATA AND ANALYTICS MANAGER

DATA SCIENCE TEAM LEADER

Role

Manages a team of analysts and data scientists

Mindset

Data Wizards' Cheerleader



HIRED BY



Languages

SQL, R, SAS, Python, Matlab, Java

Skills & Talents

- ✓ Database systems (SQL and NO SQL based)
- ✓ Leadership & project management
- ✓ Interpersonal communication
- ✓ Data mining & predictive modeling

DATABASE ADMINISTRATOR

'DATABASE CARETAKER'

Role

Ensures that the database is available to all relevant users, is performing properly and is being kept safe

Mindset

Master of Disaster Prevention



HIRED BY



SQL, Java, Ruby on Rails, XML, C#, Python

Skills & Talents

- ✓ Backup & recovery
- ✓ Data modeling and design
- ✓ Distributed Computing (Hadoop)
- ✓ Database systems (SQL and NO SQL based)
- ✓ Data security
- ✓ ERP & business knowledge

BUSINESS ANALYST

'CHANGE AGENT'

Role

Improves business processes as intermediary between business and IT

Mindset

Resilient project juggler



HIRED BY



Languages

SQL

Skills & Talents

- ✓ Basic tools (e.g. MS Office)
- ✓ Data visualization tools (e.g. Tableau)
- ✓ Conscious listening and storytelling
- ✓ Business Intelligence understanding
- ✓ Data modeling

Red Ocean vs Blue Ocean

Red Ocean	Blue Ocean
Large Job Portals	Niche Platforms & Communities
“Anonymous” Applications	Networking
Generalist	Specialist
Certificates	Unique GitHub Projects

Otras plataformas

<https://shapr.co/>

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

<https://lunchclub.com/>

<https://trycircular.com/>

BIG DATA LANDSCAPE 2017

INFRASTRUCTURE

HADOOP ON-PREMISE

cloudera Hortonworks
MAPR Pivotal
IBM InfoSphere
bluedata jethro

HADOOP IN THE CLOUD

amazon Microsoft Azure
Google Cloud Platform
IBM InfoSphere
altiscale
CenturyLink

STREAMING / IN-MEMORY

amazon databricks
confluent stream
GridGain METAMARKETS
DATATORRENT dataArtisans
ORACLE hazelcast TERRACOTTA

NOSQL DATABASES

Google Cloud Platform
ORACLE amazon dynamodb
Microsoft Azure MarkLogic
mongoDB DATASTAX
HEROSPIKE Couchbase
redislabs Influidio

NEWSQL DATABASES

SAP Clustrix
Pivotal nuodb
Cockroach Labs memsql
splice
Memorize VOLTDB
citrusdata Trafalgar
databricks paradigm4

GRAPH DBS

neo4j
IBM ORACLE
OrientDB
Cepheid

MPP DBS

TERADATA
VERTICA
NETEZZA
Celonis
kognitio
SQL dremio
InfoWorks

CLOUD EDW

amazon
Google Cloud Platform
Microsoft Azure Pivotal
snowflake
Infoworks

DATA TRANSFORMATION

talend pentaho
alteryx TRIFACTA
tamr Paxata
StreamSets UNIFI

DATA INTEGRATION

Informatica MuleSoft
snapLogic TEALUM
segment enigma
pudium aloma
xplenty ZALONI
Import Stitch

DATA GOVERNANCE

informatica
IBM skyhigh
collibra
Alation

MGMT / MONITORING

amazon New Relic
APPROXIMATIONS actifio
WAVEFRONT
unravel
splunk
pagerduty
Numerify

STORAGE

amazon
Google Cloud Platform
Microsoft Azure
ALLUXIO
CUMULUS COHO
penstest

CLUSTER SERVICES

amazon
Google Cloud Platform
Microsoft Azure
ALLUXIO
CUMULUS COHO
penstest

APP DEV

Licorband
darker
MESOSPHERE
Core OS
CRSK

CROWDSOURCING

amazon mechanical turk
upwork
WorkFusion
CrowdFlower

HARDWARE

Google TPU ARM
nervana
Graphcore
MYTHIC
NVIDIA
Movius
SCORTEX

CROSS-INFRASTRUCTURE/ANALYTICS

amazon Google Cloud Platform Microsoft IBM SAP Hewlett Packard Enterprise SAS HADOPIA vmware TIBCO TERADATA ORACLE NetApp

ANALYTICS

DATA ANALYST PLATFORMS

Microsoft pentaho alteryx
Digital Reasoning guavus AYASDI
ATTIVIO Datameer Quid
ClearStory OrigamiLogic interana
Bottlenose ARIMO ENDOR MODE

DATA SCIENCE PLATFORMS

IBM KNIME dataiku
DOMINO yhat rapidminer
CONTINUUM ANALYTICS
ALGORITHMIA Alpine
Angoss

BI PLATFORMS

Microsoft
amazon
Google Cloud Platform
looker Wave Analytics
atascade
GoodData
SIBSENSE

VISUALIZATION

tableau SAP
Google Cloud Platform
Qlik CELONIS
Periscope ZEPL
CHARTIO plotly
KODATA

VERTICAL ANALYTICS

PREDIX
Gigamon CAPE
UPTAKE
Global Insight
TACHYUS
deloitte

STATISTICAL COMPUTING

SAS
SPSS
MATLAB

DATA SERVICES

Palantir
DataKind
DataFusion
DataFusion
DataFusion

MACHINE LEARNING

Acron
Google Cloud Platform
H2O
DataRobot
ContextRelevant
VIZEN
bonsai
DataRPM
Nutanix

HORIZONTAL AI

IBM Watson Cortana
Facebook sentiment
Voyager
Affective
Coronacom
nara
OSARO
CURIOUS AI
BLUE VISION

SPEECH & NLP

Google Cloud Platform
NarrativeScience
Semafor
Indica
ARRIA
IDIBON
snips
yescop
Gridspace
SoundHound Inc.

SEARCH

elastic
Autonomy
ORACLE
ThoughtSpot
Algolia
swiftype MAANA
alphasense
Searchkick SINEOIA

LOG ANALYTICS

splunk
sumologic
loggly
woban
logz.io

SOCIAL ANALYTICS

Hootsuite
NETBASE
DATAFISHER
synthesio
bitly
predata

WEB / MOBILE / COMMERCE ANALYTICS

Google Analytics
mixpanel
sumail
retention
SIGOPT
granify
custora

APPLICATIONS - ENTERPRISE

SALES

insideSALES.COM
clari AVISO TACT
fusejmachines TROOPS

MARKETING - B2B

RADUIS App Annie
EVERSTRING Lattice
infer MINTIGO
sense tubular
DataFox ENGAGIO

MARKETING - B2C

Zeta bloomreach
blueyonder [PERSADO]
kahuna ACTIONIQ
SAILTHRU BLUECORE
quantifind
mparticle Ampero

CUSTOMER SERVICE

MEDALLIA
CLARABRIDGE
CLICKFOX
Digital Genius
AUTOMAT
mgca INTERCOM
zendesk
Gainsight
NGDATA
eppur
frame.ai

HUMAN CAPITAL

Hiive entelo
hiIQ GIGSTER
Fextia
Wade & Wendy
Custora
Stella
pymetrics

LEGAL

RAVEL
QSeal
JUDICATA
Brevia
R&SS
Casepoint

FINANCE

anaplan
Zuora
tdemark
ZEPHANA
TRADESHIFT

ENTERPRISE PRODUCTIVITY

slack
ORACLE
Lumata
Clara
talla
KASIST

BACK OFFICE AUTOMATION

HyperScience
Captricity
AppZen

SECURITY

TANUM
CYLANCE
StackPath
DARKTRACE
illumio
ThreatMetrix
DataGravity
VETRA
CiphCloud
Guardian
ANOMALI
siftscience
SIGNIFY
Recorded Future
feedzai
scure
FORTISCALE
Kalyan

APPLICATIONS - INDUSTRY

ADVERTISING

AppNexus
critico
xAd
Integral
OpenX
MOAT
drawbridge
TAPAD
DataXu
Oppier

EDUCATION

Knowledge
Clever
Gedara
kidaptive
PANORAMA
knowtop
gradscope

GOVERNMENT

Socrata
OPENGOV
mark43
FiscalNote
OpenDataSoft

FINANCE - LENDING

OnDeck Affirm
Kreditech AVANT
INSIKT
TALA MoneyLion
TrueAccord
cignifi
aire

FINANCE - INVESTING

Dataminr
KENSCH
Quantopian
NUMERA
ISSENTIUM
claritymoney
ALGORIZ
RavenPack

REAL ESTATE

Opendoor
VTS
CREDS
reonomy
COMPSTAK

INSURANCE

Metromile
Lemonade
CYENCE
Shift Technology
Tractable

HEALTHCARE

FLATIRON
HealthTap
Gingerio
COTA
AICure
imga
MAGN

LIFE SCIENCES

23andMe
colorX
zymergen
BerkeleyAI
ZEPHYRUS
HEALTH
Clear Labs
Citrine
twoAR
Atomwise
prosperra

TRANSPORTATION

UBER
TESLA
CLEARPATH
drive.ai
PTIMUS
riexor
comma.ai
netradyn

AGRICULTURE

FARMERS
FarmLogs
BLUE RIVER
mavrx
terroterra
prosperra

COMMERCE

instacart
STITCH FIX
RetailNext
HowGood

OTHER

eHarmony
stern
happn
BOKEVER
VERIDEX
duetto
KNOX
Sentry

OPEN SOURCE

FRAMEWORK

hadoop
Hadoop
Flink
YARN
TEZ
MESOS
CDAP
Spark

QUERY / DATA FLOW

SQL
presto
SLAMDATA
Google Cloud Dataflow

DATA ACCESS

cassandra
mongoDB
CouchDB
HBASE
Spanner
accumulo

COORDINATION

talend
Apache Zookeeper
Apache Ambari

STREAMING

Spark
Flink
kafka
druid
STORM

STAT TOOLS

python
ScalaLab
NumPy
SciPy

AI / MACHINE LEARNING / DEEP LEARNING

theano
TensorFlow
Caffe
CNTK
mxnet
FeatureFu
DSSTNE
mlib
DL4J
Aerosolve

SEARCH

elasticsearch
Solr
Lucene

LOG ANALYSIS

elasticsearch
kibana
logstash

VISUALIZATION

BEAKER
Rodeo

COLLABORATION

jujiter
ANACONDA

SECURITY

Apache Ranger
KNOX
Sentry

DATA SOURCES & APIS

HEALTH

JAWBONE
VALIDIC
practicefusion
fitbit GARMIN
Human API
kinsa

IOT

GE Digital
UPTAKE
ThingWorx
helium
samsara
Autodesk

FINANCIAL & ECONOMIC DATA

Bloomberg THOMSON REUTERS
S&P CAPITAL IQ CB Insights xignite
YODLEE
PREMIERE
Eagle Alpha
Stacktwits
PLAID
Thinknum

AIR / SPACE / SEA

PLANET
Airware
spire
AIRBOTICS
UNICORP
TELLUSLABS
DroneDeploy
MantaTraffic

PEOPLE / ENTITIES

axiom
Epsilon
Crimson Hexagon
BASIS
quantcast
SAFEGRAPH

LOCATION INTELLIGENCE

FOURSQUARE
Sense
esri
CARTO
Mapillary
STREETLINE

OTHER

qualtrics
DATA.GOV
dataworld
panjiva
enigma

DATA RESOURCES

INCUBATORS & SCHOOLS

PLURALSIGHT
DataCamp
Data Elite
INSIGHT
The Data Incubator

RESEARCH

facebook research
OpenAI
MIRI
Allen Institute
AI-ARTIFICIAL INTELLIGENCE

Ah, entonces tengo que sacarme
certificaciones y cursos

NO!

HAZ PROYECTOS. Son el FIN para aprender, los
cursillos y libros el MEDIO.

Un proyecto te aporta más motivación y mayor
visibilidad en tu CV