

Veri Bilimine Giriş



Veri Nedir?

Ham (işlenmemiş) gerçek enformasyon parçacığına verilen addır. Veriler ölçüm, sayım, deney, gözlem ya da araştırma yolu ile elde edilmektedir. Ölçüm ya da sayım yolu ile toplanan ve sayısal bir değer bildiren veriler nicel veriler, sayısal bir değer bildirmeyen veriler de nitel veriler olarak adlandırılmaktadır.



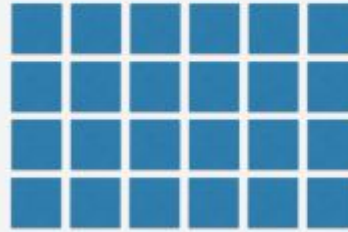
Veri Nerede Saklanır?

İki çeşit veri vardır. Birisi **Structured veri**, Yayılım grafiğine dökebileceğiniz, nicel veya nitel değerler taşıyan veri, İkincisi ise **Unstructured veri** yani nicel veya nitel veri harici, görsel, ses dosyası, email'ler, video, sosyal medya verisi vs.

Birincisi excel, veritabanları vs. yerlerde taşınabilir.

İkincisi için veri çok büyük ve anlaşılamayacağından Hard diskler, bulut ve sunucularda özellikle veri merkezlerinde saklanır. Örn. Google Data Center

Structured Data

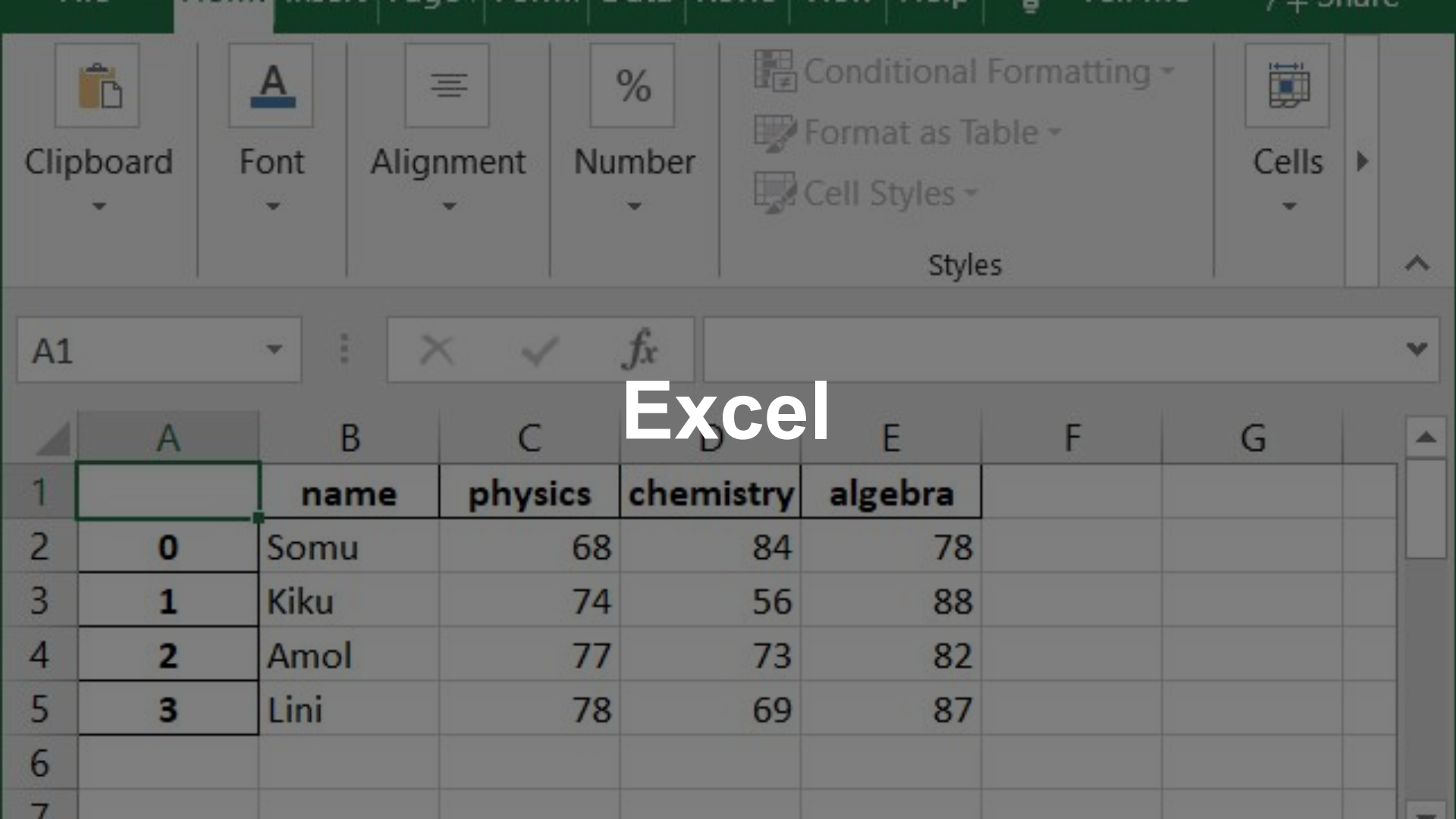


What you find in a DB
(typically)

Unstructured Data



What you find in the 'wild'
(text, images, audio, video)



A1

Excel

A

B

C

E

F

G

1

name

physics

chemistry

algebra

2

0

Somu

68

84

78

3

1

Kiku

74

56

88

4

2

Amol

77

73

82

5

3

Lini

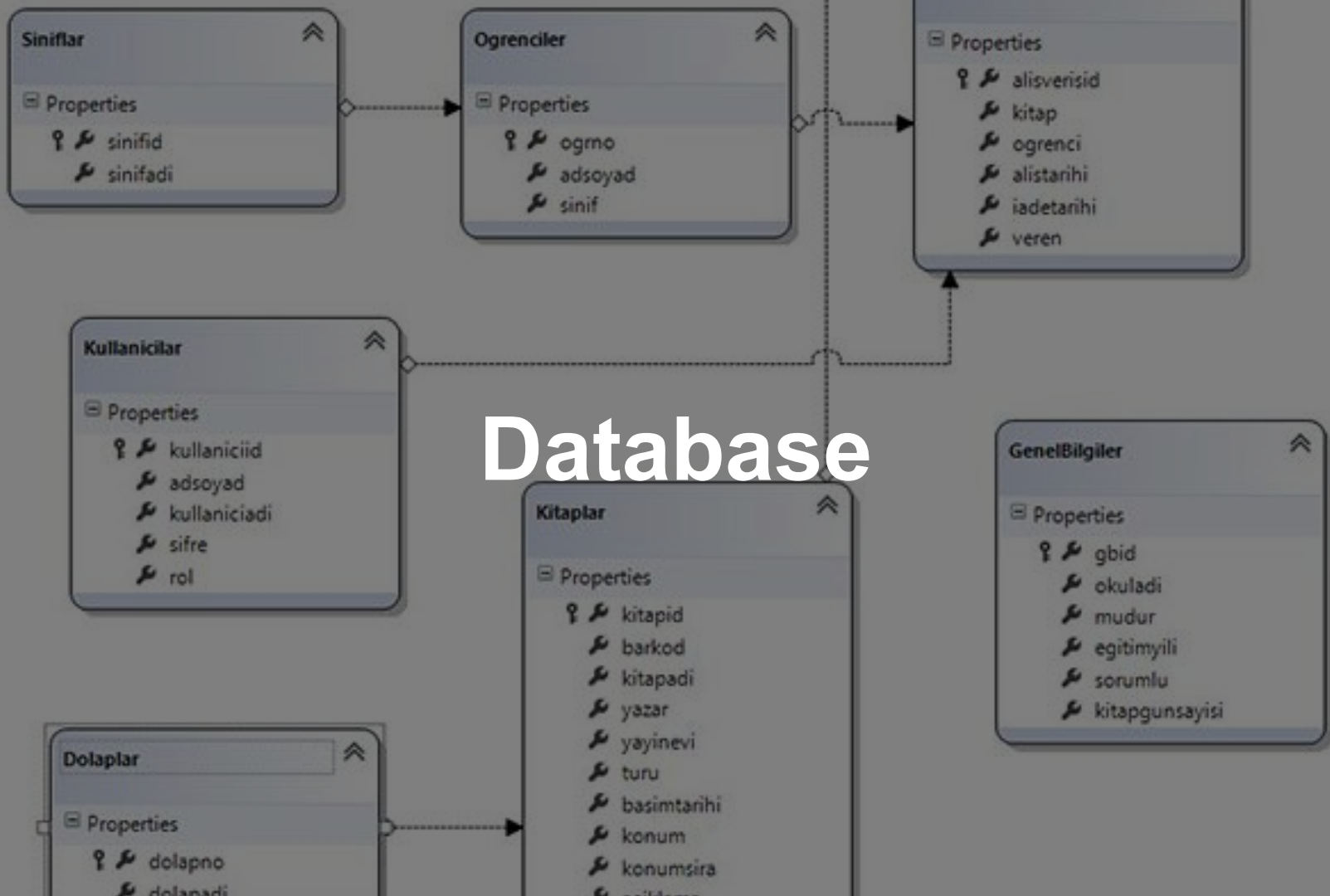
78

69

87

6

7





Data Center

Veri ne işe yarar?

Veri tek başına bir işe yaramaz. Üzerinde gözlem yapılabilir fakat bunlar yetersizdir. Veri analizi, çeşitli görselleştirmeler ve temizleme yapılması gerekir.

Bunun için biz Python programlama dilini ve Jupyter Lab çalışma ortamını kullanacağız. Alternatif olarak kullanabileceğiniz araçlar ve diller;

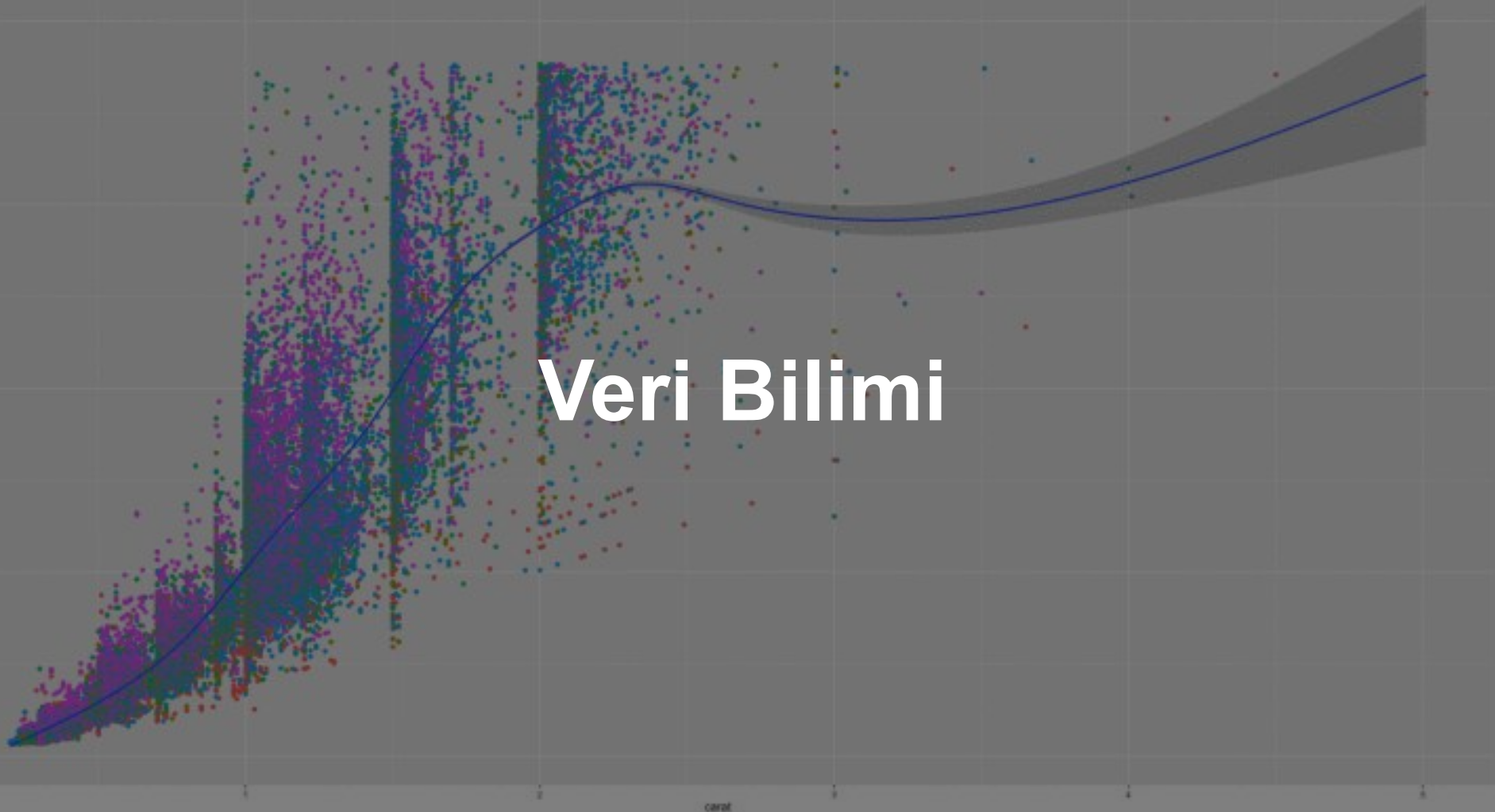
RLang, Excel, PyCharm, Colab vs.



Veri Nasıl işlenir ?

Temel olarak veriye ilk bakış insan öngörüsüyle oluşur. Daha sonra bu öngörülerin arkasına bir dayanak sağlamak için ve işin insan öngörüsünden çıkıp bilimsel bir açıklamaya dönmesi için çeşitli İstatistiksel Analiz yöntemi kurarız. Çoğunlukla Veriyi daha kolay anlamamız için görselleştirmemiz gerekir.

Veri Bilimi



Peki Veri Bilimi Nedir ?



Hiç şüphesiz , birkaç farklı şekilde zaten veri bilimini deneyimledin. Web’de arama motorunu bir bilgi için kullandığında ya da yönergeler için telefonunu kullandığında veri biliminin meyvesini yemiş bulunuyorsun. Veri bilimi birkaç yıldır bu tür çok bilindik günlük görevleri sistemin back-end tarafında çözücüsü haline gelmiştir.

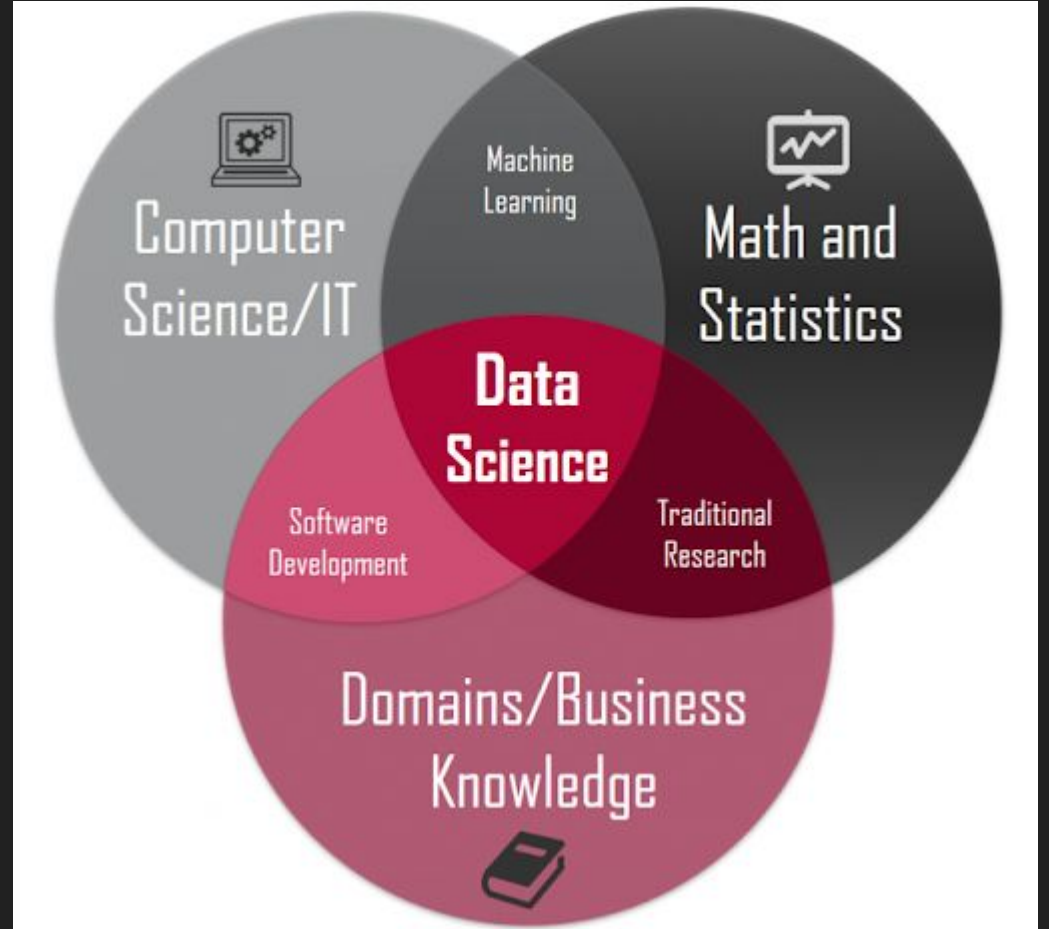
Veri Bilimci Olmak için Ne Gerekir ?

Veri bilimcilere bir zamanlar istatistikçi deniliyordu. Veri bilimi kısmen, İstatistik denince aklınıza gelenin aynısıdır.

İki meslek tabiki de aynı değil, ancak birçok veri bilimci istatistik mezunudur.

Veri bilimci olmak için az da olsa İstatistik ve Cebir bilgisi gerekir. Matematiğe sıcak yaklaşmanız gerekir.

Onun dışında İş bilgisi ve Programlama da oldukça önemli rol oynar.



Veri bilimi için gerekli bilgi ve yetenekler

Tabii ki her konuda uzman olmanız beklenmiyor fakat her konuda bilgi ve deneyim sahibi olmanız gerekli.



kaynak:vbo

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of the 21st century, requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative



PROGRAMMING & DATABASE

- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing packages, e.g., R
- ☆ Databases: SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS

COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

Veri bilimi ve Yapay Zekanın acı gerçekleri

Veri bilimci olmak basit değildir. Oldukça zordur. Şu an çoğu kursta ya da videolarda gördüğünüz gibi değildir. Çoğu kurs ve video size kod yazmayı anlatır ama teorikte çok fazla şey anlatmaz ve gerçek hayat problemlerinde zorluk çekersiniz.

Matematik ve İstatistik, özellikle İstatistiksel analiz ve betimleme çok önemlidir.

Ve daha yeni oturan bir kültür, self-taught kültürü. Bir fark yaratmadığınız sürece diplomasız veri bilimci olarak iş bulmak en azından birkaç yıl için biraz zor olabilir. Çok çabalamanız gerekecek.

Kaynak:

**Introduction To Data
Science - Springer**

Edureka

VBO

Wikipedi

Nirvacana