

Milestones in Data Science

Data Science History · Answer Key · 10 Questions

1. Who is credited with coining the term 'Data Science' in 1974 in his book 'Concise Survey of Computer Methods'?

- A) Peter Naur**
- B) John Tukey
- C) C.F. Jeff Wu
- D) William S. Cleveland

2. In which year was the term 'Big Data' first officially used in a technical context, often attributed to the report by NASA researchers Michael Cox and David Ellsworth?

- A) 1992
- B) 1997**
- C) 2001
- D) 2005

3. The term 'K-Means' for the popular clustering algorithm was first coined by which researcher in 1967?

- A) James MacQueen**
- B) Stuart Lloyd
- C) Edward Forgy
- D) Jean-Baptiste Macris

4. What was the name of the first programming language designed specifically for statistical computing, released by Bell Labs in 1976?

- A) Python
- B) S**
- C) R
- D) SAS

5. In 1996, the 'CRISP-DM' methodology was initiated; what does the acronym stand for?

- A) Cross-Industry Standard Process for Data Mining**
- B) Computational Research in Statistical Programming
- C) Coordinated Research Initiative for Science and Data Management
- D) Comprehensive Review of Integrated Statistical Data Models

6. Which seminal paper, published by Leo Breiman in 2001, introduced the 'Random Forest' algorithm to the data science community?

A) Random Forests

- B) Statistical Modeling: The Two Cultures
- C) Machine Learning Foundations
- D) Decision Tree Ensembles

7. Which company is credited with developing 'Hadoop', the framework that revolutionized big data processing, in 2006?

A) Google

B) Yahoo

- C) Amazon
- D) Microsoft

8. The first implementation of the 'Support Vector Machine' (SVM) algorithm was proposed in 1992 by which group of researchers?

A) Boser, Guyon, and Vapnik

- B) Breiman and Friedman
- C) Hastie and Tibshirani
- D) Bengio and Lecun

9. In 2008, the job title 'Data Scientist' was formally popularized by Jeff Hammerbacher and DJ Patil while they were working at which two companies?

A) Facebook and LinkedIn

- B) Google and Amazon
- C) Twitter and Uber
- D) Microsoft and Apple

10. What was the primary focus of the 'Netflix Prize', which launched in 2006 and significantly advanced collaborative filtering research?

A) Predicting movie ratings

- B) Detecting fraudulent account usage
- C) Optimizing video streaming latency
- D) Automating film metadata tagging