1. Sự khác nhau giữa Data Analyst và Data Scientist

|  |  |
| --- | --- |
| DA | DS |
| Không buộc phải biết nhiều ngôn ngữ lập trình  Cần viết báo cáo trình bày kết quả  Làm việc trên dữ liệu sẵn | Yêu cầu phải biết nhiều ngôn ngữ lập trình  Cần hiểu biết sâu rộng về toán học và khoa học máy tính, các mô hình  Làm việc với nhiều dữ liệu khác nhau cần phải lọc và khám phá tìm hiểu bộ số liệu |

1. So sánh các loại data

|  |  |  |
| --- | --- | --- |
| Chuỗi thời gian | Số liệu chéo | Số liệu hỗn hợp |
| Chuỗi số liệu được thu thập trong một thời kỳ hay một khoảng thời gian lặp lại như nhau trên cùng một không gian, một địa điểm. | Các số liệu về một hay nhiều biến được thu nhập tại cùng một thời điểm (thời kỳ) ở các khoảng không gian (địa phương, đơn vị khác nhau..) khác nhau | Các số liệu được thu nhập theo thời gian và không gian. |

3. Các bước làm dự án DA

B1: Lấy dữ liệu

B2: Làm sạch dữ liệu

B3: Biến đổi dữ liệu

B4: Phân tích dữ liệu

B5: Trình bày kết quả