



# Data Science CareerFest

**Khaleequ Ansari**  
Data Scientist



About Me

# **Khaleequ Ansari**

[linkedin.com/in/khaleequansari/](https://linkedin.com/in/khaleequansari/)

**B.Tech from IIT, Delhi**  
**3+ years experience in Data**  
**Science**  
**(Recommender systems, NLP**  
**systems)**

# Contents



## Motivation

Use cases of Data Science



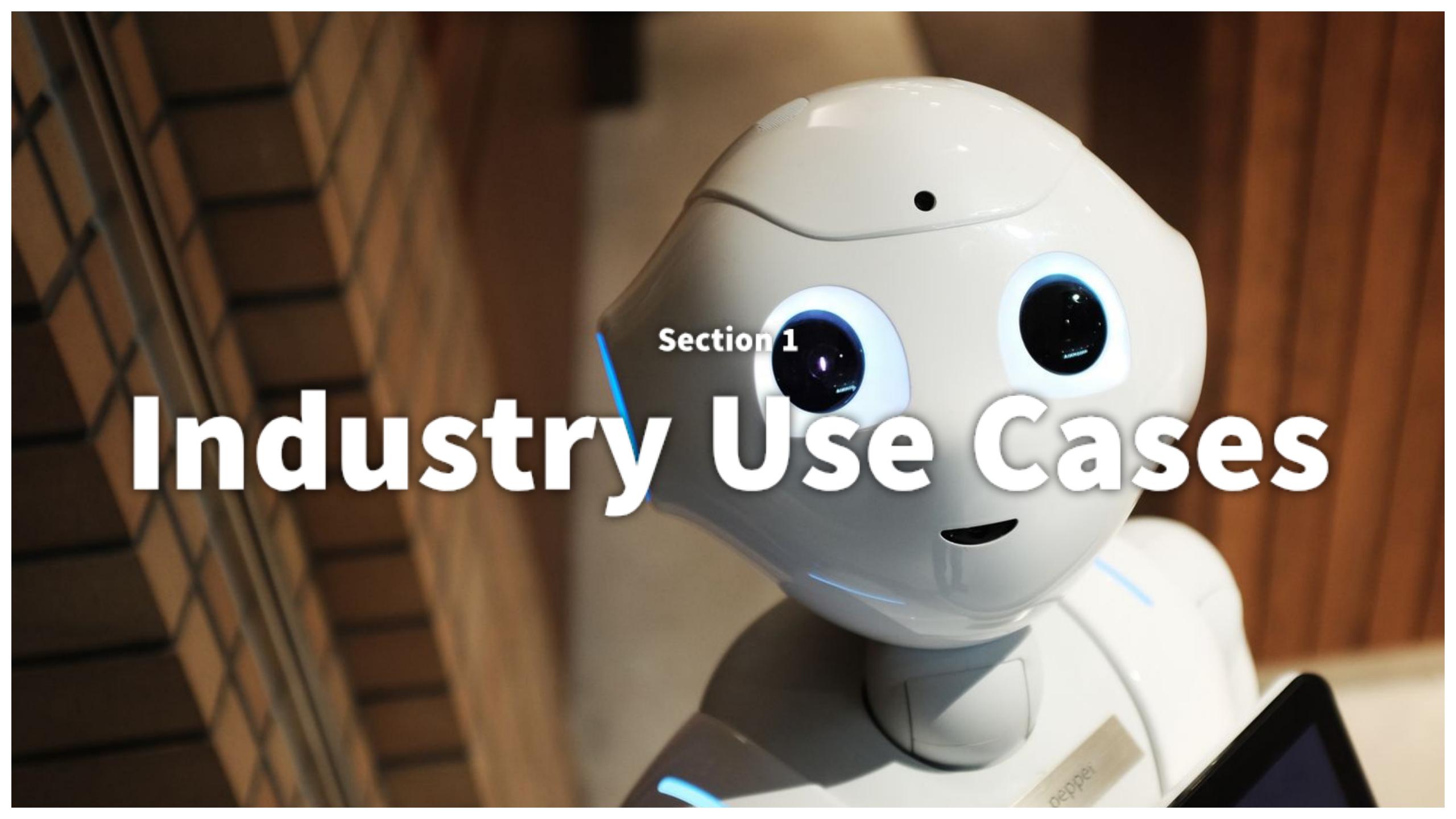
## Working Example

A sample problem and Solution



## Careers

Career choices in Data Science



# Industry Use Cases

Section 1

# Banking Industry

1 Fraud Detection

2 Customer Segmentation

3 Personalized Marketing

4 Risk Management



# Insurance Industry

1 Risk Assessment

2 Fraud Detection

3 Customer Experience

4 Customer Insights



# Social Network

1 Facial Recognition

2 Targeted Advertising

3 Textual Analysis

4 Designing AI  
applications



# Ecommerce Industry

1 Predict Trends

2 Optimise Pricing

3 Forecast Demand

4 Personalisation



# Code Examples

```
31     def __init__(self, path=None, debug=False):
32         self.file = None
33         self.fingerprints = set()
34         self.logduplicates = True
35         self.debug = debug
36         self.logger = logging.getLogger(__name__)
37         if path:
38             self.file = open(os.path.join(path,
39                             'fingerprint_log.txt'), 'w')
40             self.file.seek(0)
41             self.fingerprints.update(fp for fp in
42                                     self.read_file())
43             self.fingerprints.add('')
44         return cls(job_dir(settings['job_dir']),
45                    job_id=settings['job_id'],
46                    debug=debug)
47
48     @classmethod
49     def request_seen(self, request):
50         fp = self.request_fingerprint(request)
51         if fp in self.fingerprints:
52             return True
53         self.fingerprints.add(fp)
54         if self.file:
55             self.file.write(fp + os.linesep)
```



A professional man in a dark suit, white shirt, and blue striped tie is shown from the chest down. He is adjusting his cufflinks with his right hand while looking at his left wrist where a silver-toned watch is visible. The background is blurred, showing what appears to be an office environment with horizontal blinds.

Section 3

# Data Science Careers

## Languages

*R, SAS, Python, Matlab, SQL,  
Hive, Pig, Spark*

## Skills & Talents

- ✓ *Distributed computing*
- ✓ *Predictive modeling*
- ✓ *Story-telling and visualizing*
- ✓ *Math, Stats, Machine Learning*



# DATA SCIENTIST

'AS RARE AS UNICORNS'

## Role

*Cleans, massages and organizes  
(big) data*

## Mindset

*Curious data wizard*

HIRED BY



# DATA ANALYST

DATA DETECTIVE

## Role

*Collects, processes and performs statistical data analyses*

## Mindset

*Intuitive data junkie with high "figure-it-out" quotient*



HIRED BY

IBM hp DHL

## Languages

*R, Python, HTML, Javascript, C/C++, SQL*

## Skills & Talents

- ✓ *Spreadsheet tools (e.g. Excel)*
- ✓ *Database systems (SQL and NO SQL based)*
- ✓ *Communication & visualization*
- ✓ *Math, Stats, Machine Learning*

# 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



HIRED BY

VISA Coca-Cola logitech

## 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*

# DATA ENGINEER

SOFTWARE ENGINEERS BY TRADE

## Role

*Develops, constructs, tests and maintains architectures  
(such as databases and large-scale processing systems)*

## Mindset

*All-purpose everyman*



HIRED BY



Spotify



## Languages

*SQL, Hive, Pig, R, Matlab, SAS, SPSS, Python, Java, Ruby, C++, Perl*

## Skills & Talents

- ✓ *Database systems (SQL & NO SQL based)*
- ✓ *Data modeling & ETL tools*
- ✓ *Data APIs*
- ✓ *Data warehousing solutions*

# STATISTICIAN

'HISTORIC LEADERS OF DATA'

## Role

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

## Mindset

*Logical and enthusiastic stats genius*



HIRED BY

LinkedIn

Johnson & Johnson



PEPSICO

## 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

# DATABASE ADMINISTRATOR

'DATABASE CARETAKER'

## Role

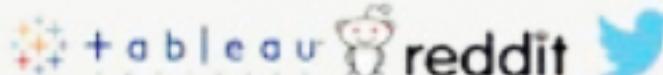
*Ensures that the database is available to all relevant users, is performing properly and is being kept safe*

## Mindset

*Master of Disaster Prevention*



HIRED BY



## Languages

*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*

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



## BUSINESS ANALYST "CHANGE AGENT"

### Role

*Improves business processes as intermediary between business and IT*

### Mindset

*Resilient project juggler*

# DATA AND ANALYTICS MANAGER

DATA SCIENCE TEAM LEADER

## Role

*Manages a team of analysts and data scientists*

## Mindset

*Data Wizards' Cheerleader*



HIRED BY

coursera

slack

MOTOROLA SOLUTIONS

## 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*

# Data Science Job Portals

- [kaggle.com/jobs](https://kaggle.com/jobs)
- [jobs.dataelixir.com](https://jobs.dataelixir.com)
- [analytictalent.com](https://analytictalent.com)
- [kdnuggets.com/jobs](https://kdnuggets.com/jobs)
- [analyticsindiajobs.com](https://analyticsindiajobs.com)
- [analyticsvidhya.com/jobs](https://analyticsvidhya.com/jobs)

# Thank You

Khaleeque Ansari

